



OPEN

## Identification of influential weather parameters and seasonal drought prediction in Bangladesh using machine learning algorithm

Md. Abdullah Al Mamun<sup>1</sup>, Mou Rani Sarker<sup>2</sup>, Md Abdur Rouf Sarkar<sup>3,4,✉</sup>, Sujit Kumar Roy<sup>5</sup>, Sheikh Arafat Islam Nihad<sup>6</sup>, Andrew M. McKenzie<sup>7</sup>, Md. Ismail Hossain<sup>1</sup> & Md. Shahjahan Kabir<sup>8</sup>

Droughts pose a severe environmental risk in countries that rely heavily on agriculture, resulting in heightened levels of concern regarding food security and livelihood enhancement. Bangladesh is highly susceptible to environmental hazards, with droughts further exacerbating the precarious situation for its 170 million inhabitants. Therefore, we are endeavouring to highlight the identification of the relative importance of climatic attributes and the estimation of the seasonal intensity and frequency of droughts in Bangladesh. With a period of forty years (1981–2020) of weather data, sophisticated machine learning (ML) methods were employed to classify 35 agroclimatic regions into dry or wet conditions using nine weather parameters, as determined by the Standardized Precipitation Evapotranspiration Index (SPEI). Out of 24 ML algorithms, the four best ML methods, ranger, bagEarth, support vector machine, and random forest (RF) have been identified for the prediction of multi-scale drought indices. The RF classifier and the Boruta algorithms shows that water balance, precipitation, maximum and minimum temperature have a higher influence on drought intensity and occurrence across Bangladesh. The trend of spatio-temporal analysis indicates, drought intensity has decreased over time, but return time has increased. There was significant variation in changing the spatial nature of drought intensity. Spatially, the drought intensity shifted from the northern to central and southern zones of Bangladesh, which had an adverse impact on crop production and the livelihood of rural and urban households. So, this precise study has important implications for the understanding of drought prediction and how to best mitigate its impacts. Additionally, the study emphasizes the need for better collaboration between relevant stakeholders, such as policymakers, researchers, communities, and local actors, to develop effective adaptation strategies and increase monitoring of weather conditions for the meticulous management of droughts in Bangladesh.

Climate change has had and continues to have catastrophic effects on humanity. Severe weather occurrences, particularly heat waves, droughts, cyclones, and heavy rain, are becoming more frequent and intense, leading to displacement, famine, and poverty<sup>1</sup>. Drought, the most frequent climate occurrence worldwide, is characterized by a shortage of precipitation which causes long-term water scarcities<sup>2–5</sup>. Droughts are one of the most expensive calamities, affecting millions of people annually and costing an estimated \$6 to \$8 billion annually<sup>6</sup>. However, the slow-onset nature of drought makes it challenging to analyze and model its spatio-temporal consequences.

Bangladesh is one of the utmost prone to natural catastrophes nations in the biosphere because of its geographical location<sup>7</sup>. Drought is a frequent natural disaster in the country. Bangladesh experienced extreme

<sup>1</sup>Agricultural Statistics Division, Bangladesh Rice Research Institute, Gazipur 1701, Bangladesh. <sup>2</sup>Sustainable Impact Platform, International Rice Research Institute, Dhaka 1213, Bangladesh. <sup>3</sup>School of Economics, Zhongnan University of Economics and Law, Wuhan 430073, China. <sup>4</sup>Agricultural Economics Division, Bangladesh Rice Research Institute, Gazipur 1701, Bangladesh. <sup>5</sup>Institute of Water and Flood Management, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh. <sup>6</sup>Plant Pathology Division, Bangladesh Rice Research Institute, Gazipur 1701, Bangladesh. <sup>7</sup>Department of Agricultural Economics and Agribusiness, The University of Arkansas, Fayetteville, AR 72701, USA. <sup>8</sup>Bangladesh Rice Research Institute, Gazipur 1701, Bangladesh. ✉email: mdrouf\_bau@yahoo.com

droughts in 1973, 1978, 1979, 1981, 1982, 1992, 1994, 1995, 2000, 2006, and 2009<sup>8</sup>. Scholars recognized that drought poses a significant risk to food security<sup>6,9,10</sup>. Climate change and scarcity of groundwater, combined with unpredictable rainfall and high temperatures, negatively impact the yields of various crops, particularly in the northwest region of Bangladesh<sup>11</sup>. Pre-kharif (mid-March to mid-May) and Rabi (mid-November to mid-March) crops are highly susceptible to drought<sup>12</sup>. Every year, droughts of varying intensities have caused damage to around 2.32 million hectares of land<sup>9</sup>. In addition to agricultural loss, drought has social and environmental consequences such as loss of livelihoods, migration, food price hikes, loss of biodiversity, disease, land degradation, and so many others<sup>1,13</sup>. Hence, drought prediction studies are necessary to reduce the adverse impacts of drought events on water resources, agriculture, energy production, ecosystems, public safety, and the economy. They are critical for sustainable resource management and preparedness in the face of a changing climate<sup>14</sup>.

Climate change adaptation and coping strategies have remained a global concern for decades. One of the key reasons for the failure of disaster risk management in climate-vulnerable countries like Bangladesh is that the government always emphasizes response and recovery over monitoring, preparedness, and mitigation. In light of this, accurate drought projections are crucial for the sustainable management of agricultural resources. The erratic and spatial nature of drought, with varying intensity and frequency<sup>10</sup>, necessitates identifying rapid, consistent, and precise prediction models to quantify drought-related risks.

Several drought indices have been established in recent decades to monitor drought on regional and global scales<sup>15–21</sup>. Among them, the standardized precipitation index (SPI)<sup>20–29</sup>, standardized precipitation evapotranspiration index (SPEI)<sup>15,18–21,27,28,30–38</sup>, and Palmer drought severity index (PDSI)<sup>39,40</sup> were widely used. Recently developed SPEI<sup>15</sup> has the advantage of determining many types of drought<sup>41</sup>. The SPEI takes into account both the multi-scalar properties and straightforward computation of the SPI and the PDSI's sensitivity to shifts in evaporation demand; hence, broadly acceptable for monitoring and analyzing drought characteristics<sup>42</sup>.

Sustainable water management requires a reliable data-driven drought prediction model<sup>43,44</sup>. Traditional stochastic techniques, such as the autoregressive integrated moving average (ARIMA) and seasonal autoregressive moving average (SARIMA) models, were the most widely used for predicting droughts<sup>45,46</sup>. Recent applications of machine learning (ML) models offer the advantage of being more adaptable and robust for drought prediction<sup>44,47–52</sup>. ML models better capture complicated relationships between variables, handling nonlinearity and temporal dependencies. Additionally, ML models can be easily updated with new data, making them suitable for dynamically changing environments<sup>53,54</sup>. Several ML models, such as artificial neural networks (ANN), Fuzzy Logic (FL), support vector regression (SVR), random forests (RF), relevance vector machine (RVM), genetic programming (GP), and extreme learning machine (ELM) have been used in complex modeling interactions<sup>46,55–57</sup>. However, because of regional variability, no generalized or ideal model is acceptable for all climates situations<sup>13</sup>; rather, there is a risk of misleading model development<sup>10,58</sup>.

In Bangladesh, very little research has been done using ML methods<sup>13,59,60</sup>. All research was one or two region-specific<sup>61</sup>, and the development of ML models for drought forecasting on a more disaggregate regional scale has yet to unfold. Besides, researchers did not identify the relative importance of climatic attributes for drought assessment. The novelty of this study is that it fills these gaps by developing the best ML models for SPEI forecasting at multiple time scales and drought intensity mapping for Bangladesh. Specifically, the current study predicts SPEIs for 35 meteorological stations using 24 ML models. Then the deployed models' performance was evaluated to select the best drought forecasting features, and finally the spatio-temporal pattern of seasonal drought intensity and frequency was estimated for meteorological research stations across Bangladesh.

The weather patterns in Bangladesh are undergoing significant transformations due to its proximity to the equator and the rising global temperatures<sup>62,63</sup>. As a consequence of these changes, the country has been confronted with severe weather fluctuations, including frequent flooding and other calamities. In regions of Bangladesh where drought is a persistent concern, experts have observed an increase in the occurrence of droughts, attributable to alterations in temperature and precipitation patterns<sup>64</sup>. The impact of climate change is anticipated to result in more substantial economic losses from droughts<sup>65</sup>, affecting water resources and contributing to water scarcity<sup>66</sup>. These adverse consequences underscore the need for the development of robust forecasting and monitoring models for drought, enabling the timely formulation of strategies to manage drought-related risks<sup>67</sup>.

Effective drought forecasting is an indispensable component of drought management. Inadequate forecasting can lead to suboptimal management practices and potential harm to the environment. Consequently, there is a pressing demand for rapid, reliable, and accurate models for drought prediction that can furnish quantitative insights into impending drought-related threats. These models leverage the appropriate combination of input variables or drought indices to deliver precise drought forecasts<sup>68</sup>.

By exploring the interplay between climatic variables and machine learning models<sup>69</sup>, we aim to uncover the optimal combination that can provide actionable insights and early warning systems for mitigating the far-reaching impacts of droughts in the region. This research represents a vital step towards harnessing technology and data-driven approaches to address the pressing challenges posed by seasonal droughts, offering the potential to safeguard agricultural practices in Bangladesh. Our research holds significant implications for science, policy, and practice. Scientifically, it showcases the efficacy of machine learning methods in drought prediction and underscores the significance of integrating weather parameters in drought analysis. From a policy perspective, the study underscores the need for enhanced collaboration among stakeholders, including policymakers, researchers, communities, and local actors, to formulate effective adaptation strategies. In practical terms, the findings highlight the crucial importance of heightened weather condition monitoring in Bangladesh to mitigate the adverse effects of droughts on crop production and livelihoods.

The article is organized as follows: "[Materials and methods](#)" provides information on the study area, the utilized data, and the empirical settings. "[Results](#)" presents the study findings. A detailed discussion of the results is found in "[Discussion](#)", with a summary of conclusions and recommendations in "[Conclusions and policy recommendations](#)".

## Materials and methods

### Study area

Bangladesh, located in South Asia, stands out for its distinctive geographical and environmental features. Situated between latitudes 20°34' and 26°38' N and longitudes 88°01' and 92°41' E (Fig. 1), it shares borders with India to the west, north, and east, and Myanmar (Burma) to the southeast<sup>70</sup>. To the south, the Bay of Bengal forms a natural boundary. This densely populated nation boasts a complex landscape with expansive riverine systems, fertile alluvial plains, and the renowned Sundarbans, the world's largest river delta. Its predominantly low-lying terrain renders it susceptible to flooding, storm surges, and monsoon rains, with intermittent droughts affecting certain regions. Featuring a tropical climate, the country experiences average temperatures ranging from 12.8 to 31.1 °C. The annual rainfall varies from 1700 mm in the northwest to over 5000 mm in the southeastern region<sup>13</sup>, establishing Bangladesh as one of the world's wettest countries. The climate features distinct wet and dry seasons, profoundly impacting agriculture, the economy, and the predominantly agrarian livelihoods of its people. Bangladesh's unique geographical and climatic conditions have made it a focal point for research in areas such as climate change, agriculture, water resource management, and disaster preparedness, underlining its critical importance due to its vulnerability to environmental challenges and the potential for innovative solutions to enhance the well-being of its population.

### Data use

Climate records at the daily timescale from 35 meteorological stations were collected by the Bangladesh Meteorological Department (BMD) over the past 40 years, from 1981 to 2020 (Fig. 1). The climate variables were daily rainfall amount (mm), maximum temperature (°C), minimum temperature (°C), mean temperature (°C), sunshine hours (h), wind speed (ms<sup>-1</sup>), and relative humidity (%). In addition, potential evapotranspiration (PET) at the monthly timescale was calculated from the aforementioned climate variables. The Food and Agricultural Organization (FAO) recommends the Penman–Monteith (PM) equation<sup>71</sup> as the single standard technique for calculating reference evapotranspiration (ET<sub>0</sub>), and it has been effectively utilized in Bangladesh. It integrates physiological and meteorological attributes and has been widely used around the world because of its intrinsic rationality and reliability<sup>72</sup>. Hence, the PM equation based on the weather parameters was utilized to compute the monthly ET<sub>0</sub> over the research locations.

### Model selection process for data analysis

In this research, twenty-four (24) machine learning models were constructed to predict the Standardized Precipitation Evapotranspiration Index (SPEI) in various timescales, including 1-, 3-, 6-, and 12-month periods. The methodology, as illustrated in Fig. 2, encompassed the subsequent procedural phases:

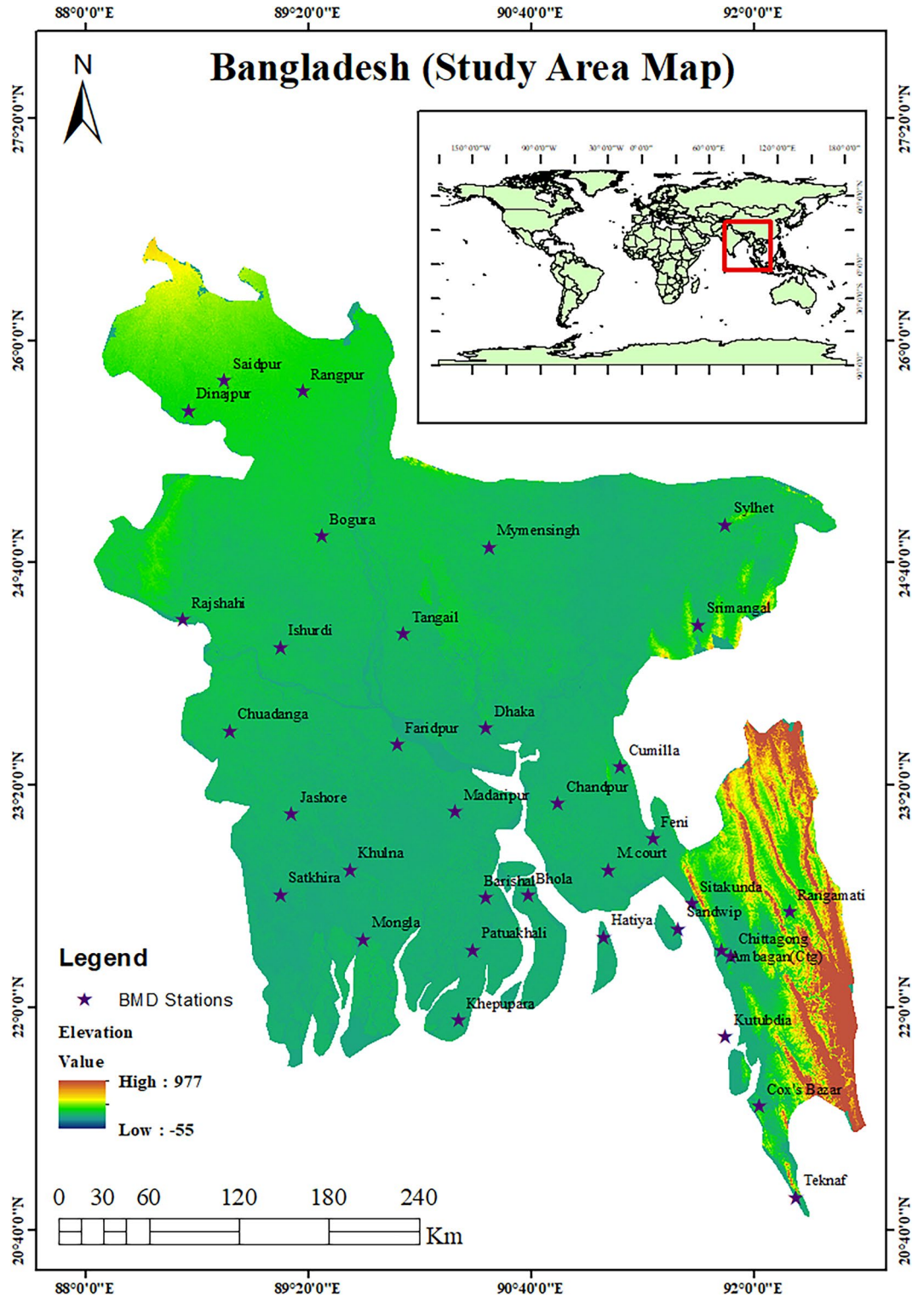
- (a) Data collection and preprocessing: The SPEI database was generated by utilizing meteorological variables from the SPEI computation model.
- (b) Feature selection: The optimal features for classifying drought conditions were identified through the application of correlation and random forest classifier methods.
- (c) Model selection and cross-validation: To implement machine learning algorithms, the dataset was randomly divided into training (80%) and testing (20%) sets, and all data points from each research station were imported into the R programming environment. Machine learning algorithms were employed on the training datasets, the models were validated using the testing datasets, and SPEI predictions were generated for various time scales. Each meteorological station's resulting output values were ranked to determine the most suitable machine learning models.
- (d) Model evaluation: Models were evaluated using appropriate performance metrics, including RMSE, MAE, and R<sup>2</sup>.
- (e) Model comparison: We compare the performance of different models and select the one that best aligns with our research objectives and provides the most accurate results.
- (f) Validation and robustness testing: The selected model was further validated using different subsets of the data to assess its robustness.
- (g) Model output and visualization: We focus on the presentation and interpretation of the model output through a robust visualization process.

### Standardized precipitation evapotranspiration index (SPEI)

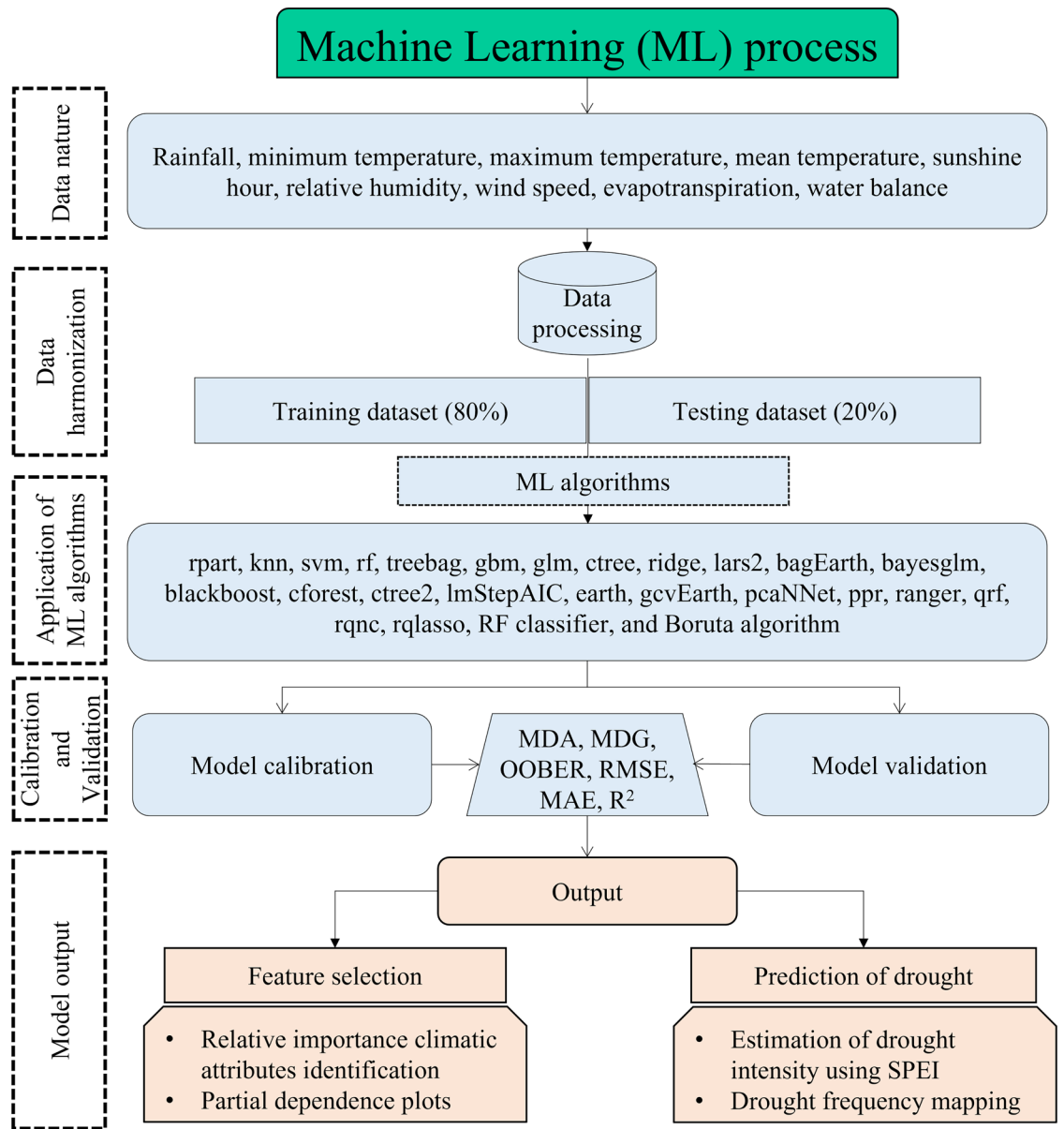
SPEI is a widely used technique for measuring drought dynamics over multiple time frames<sup>15</sup>. The SPEI is derived from precipitation and temperature data using a simple water balance to measure the effects of surface evaporation caused by increasing global temperatures<sup>73</sup>. It has an advantage over the SPI, because it combines rainfall and temperature into its computations, whereas SPI only uses rainfall<sup>22</sup>. The calculation of SPEI is based on the original SPI calculation procedure and hence uses the same index categorization criteria<sup>15</sup>.

The initial step in computing the SPEI is to determine the monthly water balance (D<sub>i</sub>), which is the difference between the precipitation (P<sub>i</sub>) and potential evapotranspiration (PET<sub>i</sub>). Afterwards, these values are combined at the desired timescales<sup>74</sup> as:

$$D_i = P_i - PET_i \quad (1)$$



**Figure 1.** A map illustrating the locations of the examined meteorological stations in Bangladesh. The authors used ArcGIS 10.8 (<https://www.arcgis.com/index.html>) to generate the map, employing the administrative shapefile of Bangladesh in the process. Shapefile republished from the Bangladesh Agricultural Research Council (BARC) database (<http://maps.barcapps.gov.bd/index.php>) under a CC BY license, with permission from Computer and GIS unit, BARC, original copyright 2014.



**Figure 2.** Conceptual framework of prediction of SPEI by ML algorithms for the study.

$$D_n^k = \sum_{i=0}^{k-1} (P_{n-i} - PET_{n-i}), n \geq k \tag{2}$$

where,  $k=1, 3, 6,$  and  $12$  for  $SPEI=1, 3, 6,$  and  $12$  is the aggregation timescale, and  $n$  is the  $n$ th month. Using the log-logistic probability distribution, the  $D$ -series was fitted. The cumulative distribution function  $F(x)$ <sup>75</sup> can be expressed as follows:

$$F(x) = \left[ 1 + \left( \frac{\alpha}{x - \gamma} \right)^\beta \right]^{-1} k \tag{3}$$

where, the parameters of scale, shape, and location are written as  $\alpha, \beta,$  and  $\gamma,$  respectively. The  $SPEI$  value is derived from the standard value provided:

$$SPEI = W - \frac{c_0 + c_1 W + c_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \tag{4}$$

$$W = \sqrt{-2\ln(P)} \text{ for } P \leq 0.5 \tag{5}$$

where  $P$  indicates the likelihood of exceeding a certain  $D$  value, and  $F(D)$  represents the cumulative distribution of  $D$ . When  $P$  is greater than  $0.5,$  it is replaced with the non-exceed likelihood ( $F(D) = 1 - P$ ) and the direction

of the derived SPEI is inverted. The constants are  $c_0 = 2.5155$ ,  $c_1 = 0.8028$ ,  $c_2 = 0.0103$ ,  $d_1 = 1.4328$ ,  $d_2 = 0.1892$  and  $d_3 = 0.0013$ <sup>43</sup>.

In this study, we estimated SPEI at time scales of one month (SPEI1), three months (SPEI3), six months (SPEI6), and a year (SPEI12). These estimates were used to measure the impact of precipitation deficits in the short term on agricultural drought. According to the SPEI classification criteria, the value of  $\text{SPEI} \geq 0$  indicates no drought,  $-1.0 < \text{SPEI} < 0$  indicates mild drought,  $-1.5 < \text{SPEI} \leq -1.0$  indicates moderate drought,  $-2.0 < \text{SPEI} \leq -1.5$  indicates severe drought and  $\text{SPEI} \leq -2.0$  indicates extreme drought<sup>76</sup>. The greater the value of the SPEI in the negative, the more severe the drought.

We also estimated the severity of drought. A drought event's duration ( $m$ ) equals the number of months between its start (included) and end month (not included). The absolute value of the total of all SPEI values during a drought event is known as severity ( $S_e$ ). A drought event's intensity ( $DI_e$ ) is defined as severity divided by duration<sup>77</sup>. The greater the  $DI_e$  number, the more severe the drought. The formulae are as follows:

$$S_e = \left| \sum_{j=1}^m \text{Index}_j \right|_e \quad (6)$$

$$DI_e = \frac{S_e}{m} \quad (7)$$

where,  $e$ ,  $j$ ,  $\text{Index}_j$ ,  $m$ ,  $S_e$ , and  $DI_e$  are the drought event, month, SPEI value in month  $j$ , duration, severity, and intensity of a drought event  $e$ , respectively.

#### Best feature selection criteria

Feature selection is a widely used process of selecting the best features that can significantly influence the predicted outcomes, increasing model performance and accuracy, and reducing running time<sup>78,79</sup>. We considered two different feature selection approaches, random forest (RF), and the Boruta algorithm, to identify the most significant weather variables that affect SPEIs. The Boruta and *Caret* packages were used for feature selection in RStudio software. A brief overview of these two techniques is provided here:

**Random forest.** RF employed Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) to select variables<sup>80</sup>. When a variable is left out of the model, the MDA value represents how much precision is compromised. The more accuracy lost the more importance of the variable for successful classification. The MDG measures the contribution of each variable to the homogeneity of the random forest's nodes and leaves. The greater the MDG score, the greater the significance of the variable in the model<sup>81</sup>. Additionally, we used dichotomy method as a rapid variable screening technique. Time series of SPEI and other related weather variables were used to find a proficient and robust estimation of the best classifier. As a rule, the majority of scholars employed permuting out-of-bag (OOB henceforth) error or impurity to evaluate the significance of a single variable<sup>82–84</sup>, whereas we employed both. The OOB error is a bootstrap aggregation-based approach for assessing the prediction error of random forests, boosted decision trees, and other machine learning models<sup>85</sup>.

**Boruta algorithm.** The Boruta algorithm was introduced by Kursa and Rudnicki<sup>86</sup>. This technique attempts to reduce misleading outcomes due to correlations and random fluctuations by introducing more randomization and collecting results from the entire set of randomized samples. The relative importance of the climatic variables was identified for the estimate approach by following the steps outlined in Ebrahimi-Khusfi<sup>87</sup>, and Kursa and Rudnicki<sup>86</sup>.

#### Machine learning algorithms

We considered analyzing 24 distinct machine learning algorithms from various ML fields to determine the correlation between drought prediction and the weather attributes. The prediction of multiscale SPEI1, SPEI3, SPEI6, and SPEI12 considered tree-based algorithms, regression, and classification models. We used multiple predictive modeling techniques employing a variable selection algorithm. These methods included linear least squares models and penalized linear, additive, and recursive partitioning models, all implemented with R programming code for variable selection and prediction (Table 1).

The study measured the performance of each ML algorithm independently. ML algorithms employ various statistical, probabilistic, and optimization methods to extract useful patterns from large and complex datasets that are unstructured and derived from past experiences<sup>61</sup>. Time series data is characterized by a sequential order, where each observation is influenced by the preceding observations. Applying traditional cross-validation techniques to time series data can introduce a significant source of bias since it violates the temporal structure of the data. To obtain robust and reliable estimates of a model's performance, assess its generalization capabilities, and make informed decisions in model selection, using tenfold cross-validation with five repetitions, the algorithms were trained and compared. It helps reduce the impact of random variations, provides more stable performance metrics, and aligns with established practices in the field of machine learning and data analysis. All predictive models were trained utilizing the *Caret* package's interface for the train function<sup>88</sup>. The train function generates the parameter tuning by determining the values that maximize root-mean-square error (RMSE) accuracy. The data were divided into training (80%) and test (20%) sets. The ML function determines the optimal subset of predictors for the best accurate model. Finally, the study rated lists of predictors from each training approach for the final models.

Sl. no	Method	Description	Type	R packages	Reference
1	rpart	CART	Regression, classification	Rpart	Breiman and Freedman <sup>122</sup> , Breiman and Ihaka <sup>123</sup> , Therneau et al. <sup>124</sup>
2	knn	k-nearest neighbors	Regression, classification	Kknn	Uddin et al. <sup>125</sup>
3	svm	Support vector machines with radial basis function kernel	Regression, classification	Kernlab	Noble <sup>126</sup>
4	rf	Random forest	Regression, classification	randomForest	Breiman <sup>80</sup>
5	treemap	Bagged CART	Regression, classification	ipred, plyr, e1071	Breiman <sup>80</sup> , Kober et al. <sup>127</sup>
6	gbm	Stochastic gradient boosting	Regression, classification	gbm, plyr	Freedman <sup>128</sup> , Guelman <sup>129</sup> , Ridgeway <sup>130</sup>
7	glm	Generalized linear model	Regression, classification	MASS	Annette J. Dobson <sup>131</sup>
8	ctree	Conditional inference tree	Regression, classification	Party	Hothorn et al. <sup>132</sup>
9	ridge	Ridge regression	Regression	Elasticnet	Seegerist <sup>133</sup> , Zou and Hastie <sup>134</sup>
10	lars2	Least angle regression	Regression	Lars	Efron et al. <sup>135</sup>
11	bagEarth	Bagged MARS	Regression, classification	Earth	Max Kuhn et al. <sup>88</sup>
12	bayesglm	Bayesian generalized linear model	Regression, classification	Arm	Dey et al. <sup>136</sup>
13	blackboost	Boosted tree	Regression, classification	Party, mboost, plyr, partykit	Chen <sup>137</sup>
14	cforest	Conditional inference random forest	Regression, classification	Party	Levshina <sup>138</sup>
15	ctree2	Conditional inference tree	Regression, classification	Party	Hothorn et al. <sup>132</sup> , Sarda-Espinosa et al. <sup>139</sup>
16	lmStepAIC	Linear regression with stepwise selection	Regression	MASS	Olusegun et al. <sup>140</sup>
17	earth	Multivariate adaptive regression spline (MARS)	Regression, classification	Earth	Friedman and Roosen <sup>141</sup> , Milborrow et al. <sup>142</sup>
18	gcvEarth	MARS generalized cross validation (GCV) penalty per knot	Regression, classification	Earth	Milborrow et al. <sup>142</sup>
19	pcaNNet	Neural networks with a principal component step	Regression, Classification	Nnet	Ripley <sup>143</sup>
20	ppr	Projection pursuit regression	Regression	Stats	Farikha et al. <sup>144</sup>
21	ranger	Random forest	Regression, classification	e1071, ranger, dplyr	Wright et al. <sup>145</sup>
22	qrf	Quantile random forest	Regression	quantregForest	Li and Peng <sup>146</sup>
23	rqnc	Non-convex penalized quantile regression	Regression	rqPen	Bello et al. <sup>147</sup> , Ma et al. <sup>148</sup>
24	rqlasso	Quantile regression with LASSO penalty	Regression	rqPen	Ciner et al. <sup>149</sup>

**Table 1.** List of 24 machine learning methods and their characterizations.

#### Model evaluation metrics

Model validation is a necessary step of ML modeling for evaluating the accuracy and reliability of models. Scholars employed various statistical metrics for this purpose<sup>89–91</sup>. We used RMSE, MAE, and R<sup>2</sup> to evaluate the performance of the constructed models. The statistical evaluation metrics are the following for all parameters:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{obs} - Y_{pred})^2}{N}} \quad (8)$$

$$MAE = \frac{\sum_{i=1}^N |Y_{obs} - Y_{pred}|}{N} \quad (9)$$

$$R^2 = \frac{\sum_{i=1}^N (Y_{obs} - \bar{Y}_{obs})(Y_{pred} - \bar{Y}_{pred})}{\sqrt{\sum_{i=1}^N (Y_{obs} - \bar{Y}_{obs})^2} \sqrt{\sum_{i=1}^N (Y_{pred} - \bar{Y}_{pred})^2}} \quad (10)$$

where,  $Y_{obs}$  and  $Y_{pred}$  indicates the actual and predicted dependent variable, respectively, with N denoting the number of observations. As a general rule, models with a lower RMSE, MAE, and a larger R<sup>2</sup> during testing were deemed more accurate for accuracy of good prediction model.

#### Partial dependence plots (PDPs) method

The model-independent method is based on determining the "flatness" of the PDPs of each feature. PDPs assist in visualizing the influence of low cardinality feature space subsets on the estimated prediction surface, such as main effects and two/three-ways interaction effects. The PDPs provides model-independent interpretations and can be developed via a supervised machine learning approach. We train a projection pursuit regression (PPR) model and use the pdp package to generate PDPs for each feature<sup>92</sup>. The PDPs can be misled in the presence of substantial interactions<sup>93</sup>. To solve this issue, Goldstein et al.<sup>93</sup> developed individual conditional expectation (ICE) charts, which are available in the R programming package 'ICEbox'. The ICE plots display the estimated

association between the response and a predictor of relevance for each observation. Consequently, the PDP for an activity predictor may be computed by averaging the proper ICE curves' overall data.

## Results

### Correlation analysis

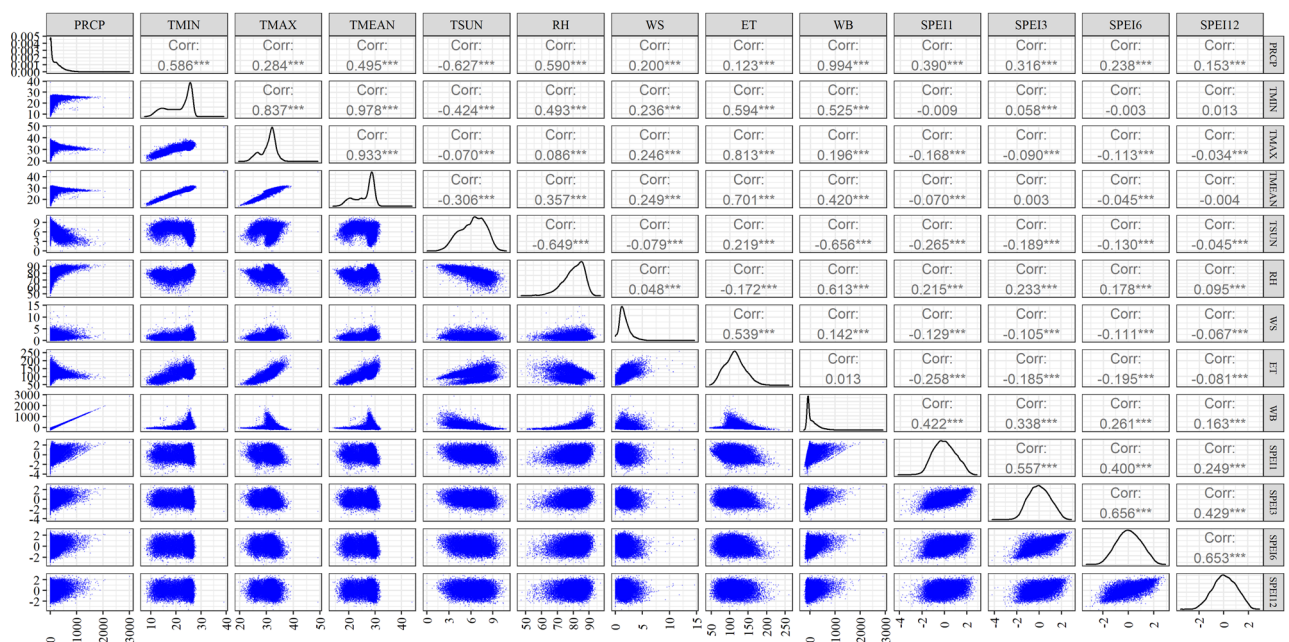
In this study, the meteorological indices, standardized precipitation evapotranspiration index at multiple time scales (SPEI1, SPEI3, SPEI6, and SPEI12), were chosen to assess the drought conditions of Bangladesh. The study aimed to explore their associations with nine distinct weather parameters: precipitation (PRCP), minimum temperature (TMIN), maximum temperature (TMAX), average temperature (TMEAN), total sunshine (TSUN), relative humidity (RH), wind speed (WS), evapotranspiration (ET), and water balance (WB) (Fig. 3). Regarding SPEI1, the correlation coefficients revealed statistically significant ( $p < 0.05$ ) and positive associations with rainfall (0.390), relative humidity (0.215), and water balance (0.422). Conversely, there were statistically significant negative correlations with maximum temperature ( $-0.168$ ), mean temperature ( $-0.070$ ), total sunshine hour ( $-0.265$ ), wind speed ( $-0.129$ ), and evapotranspiration ( $-0.258$ ). The correlation results revealed that meteorological indices at the 3-month time scale (SPEI3) exhibited a nearly identical relationship like SPEI1 with meteorological variables. Specifically, rainfall, minimum temperature, relative humidity, and water balance displayed significant positive associations with SPEI3, with correlation coefficients of 0.316, 0.058, 0.233, and 0.338, respectively. On the other hand, there were statistically significant negative correlations with maximum temperature ( $-0.090$ ), total sunshine hour ( $-0.189$ ), wind speed ( $-0.105$ ), and evapotranspiration ( $-0.185$ ).

Also, for both SPEI6 and SPEI12, the correlation analysis reveals significant ( $p < 0.05$ ) positive associations with rainfall (0.238 and 0.153), relative humidity (0.178 and 0.095), and water balance (0.261 and 0.163) (Fig. 3). In the case of SPEI6 and SPEI12, it's noteworthy that a negative and statistically significant relationship was observed with certain meteorological parameters. Specifically, maximum temperature exhibited negative correlations of  $-0.113$  for SPEI6 and  $-0.034$  for SPEI12. Likewise, total sunshine hour displayed negative correlations of  $-0.130$  for SPEI6 and  $-0.045$  for SPEI12. Additionally, wind speed showed negative correlations of  $-0.111$  for SPEI6 and  $-0.067$  for SPEI12. Furthermore, evapotranspiration had particularly significant negative correlations, with values of  $-0.195$  for SPEI6 and  $-0.081$  for SPEI12. These results indicate that as SPEI6 and SPEI12 values decreased, these meteorological parameters tended to increase, and the relationships were statistically significant.

### Identification of best climatic attributes for different SPEIs

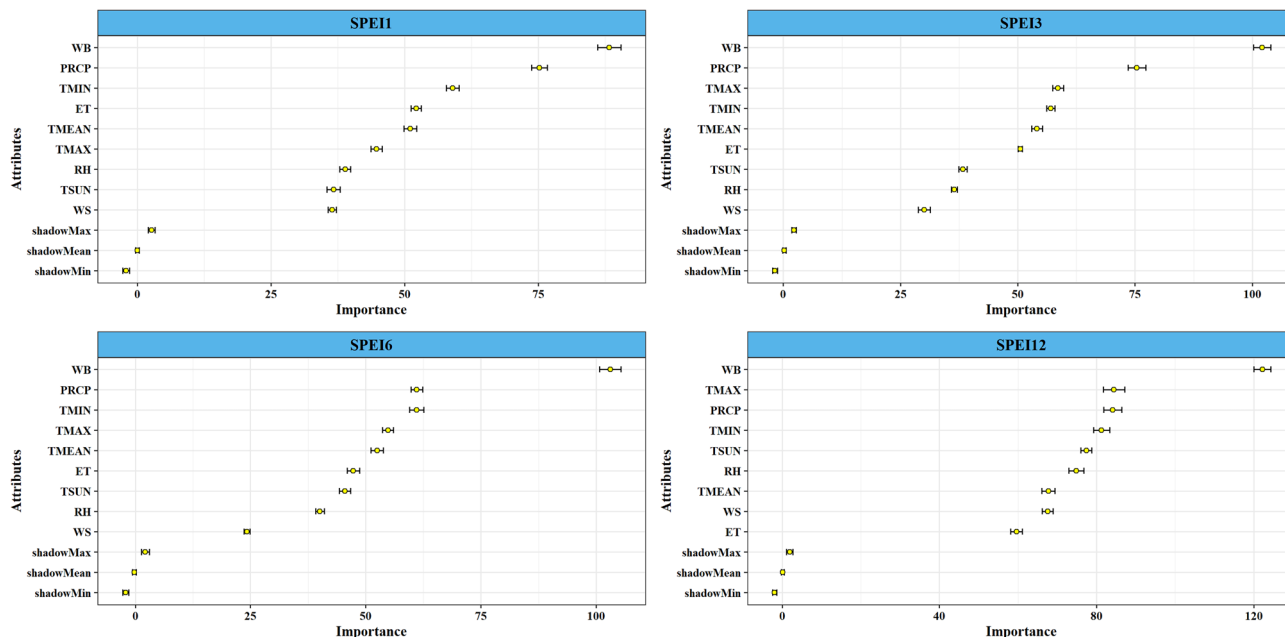
Based on the results depicted in Fig. 4, the critical variables for the SPEI1 time scale were identified as WB, PRCP, TMIN, and ET. For the SPEI3, SPEI6, and SPEI12 time scales, the most significant variables were WB, PRCP, TMAX, and TMIN. Consequently, WB, PRCP, TMAX, and TMIN emerged as the predominant factors influencing the construction of drought prediction models using machine learning.

The random forest classifier algorithms tuned using cross-validation ten folds and five repeats were summarized, and the performance of the RF classifier was presented in Table 2. The best three contributors for the SPEI1 model were WB, PRCP, and TMIN with the highest percentage values, and the overall OOB error rate for SPEI1 model was 17.77%. However, the worse contributor was identified as WS, RH, and TSUN, getting the lowest percentage among the variable for SPEI1. We found that the SPEI3 model had the same contributors



**Figure 3.** Correlation coefficients among the weather parameters and SPEI's values.





**Figure 4.** Best feature combination of predictor variables based on the Boruta algorithm.

as the SPEI1 model. The best predictor for SPEI6 was WB, followed by TMIN, PRCP, and TMAX. The worse contributor was WS, which had the lowest percentage value of MDA and MDG, but the OOB error rate was 23.05%. The annual time scale (SPEI12) also has a vital role in identifying and predicting drought. The best and most significant contributors for SPEI12 were WB, TMIN, and TMAX, and the OOB error rate was low at 6.59%. Thus, the findings indicated that WB, PRCP, TMAX, and TMIN were the most significant contributors to drought model prediction across different time scales of Bangladesh.

### Performance evaluation of ML models for different SPEIs during the training phase

To assess the precision and performance of the model, we utilized metrics including mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination ( $R^2$ ). In our interpretation, a well-performing model is characterized by lower MAE and RMSE values and a higher  $R^2$  value. For predicting SPEI1 (Standardized precipitation evapotranspiration index at a 1-month time scale), the ranger model stood out as the most accurate, boasting an impressive  $R^2$  value of 0.689, indicating its substantial explanatory power (Fig. A1a). Additionally, it displayed relatively lower RMSE (0.547) and MAE (0.417) values, reflecting close alignment with actual data. The rf, svm, and cforest models also performed well, securing the second, third, and fourth positions, respectively, in SPEI1 prediction accuracy. Conversely, the CART and lars2 models exhibited lower accuracy in predicting SPEI1.

According to Fig. A1b, evaluating the performance of machine learning models for predicting the 3-month time scale drought (SPEI3), the ranger model emerged as the top performer with an  $R^2$  of 0.602, RSME = 0.600, and MAE = 0.454, indicating robust predictive capabilities. The rf model closely followed, achieving an  $R^2$  of

Attributes	SPEI1			SPEI3			SPEI6			SPEI12		
	MDA (%)	MDG (%)	OOB error rate (%)	MDA (%)	MDG (%)	OOB error rate (%)	MDA (%)	MDG (%)	OOB error rate (%)	MDA (%)	MDG (%)	OOB error rate (%)
PRCP	13.65	15.24	17.77	13.76	15.63	19.19	10.84	14.44	23.05	9.90	17.02	6.59
TMIN	12.67	10.85		11.90	10.32		12.58	11.70		13.40	9.19	
TMAX	9.15	9.10		11.46	9.62		10.75	9.70		12.97	9.55	
TMEAN	10.03	9.64		10.72	9.43		10.39	10.13		9.33	6.91	
TSUN	7.53	7.35		7.38	7.80		10.00	8.80		10.61	8.42	
RH	8.86	6.88		7.57	8.09		8.75	9.43		10.07	10.45	
WS	9.01	6.29		7.13	6.57		5.98	7.39		9.44	7.56	
ET	11.04	9.58		10.28	8.50		9.76	8.71		7.36	8.72	
WB	18.06	25.07	19.81	24.03	20.94	19.70	16.91	22.18				

**Table 2.** Performance of RF classifier model with different SPEI time scales derived from different feature combinations. MDA mean decrease accuracy, MDG mean decrease Gini, OOB out-of-bag estimate of error rate.

0.602, RSME = 0.601, and MAE = 0.456. The qrf and cforest models secured the third and fourth positions with  $R^2$  values of 0.598 and 0.593, respectively. In contrast, the lars2 and rqlasso models exhibited lower accuracy in predicting SPEI3. The analysis was extended to predict SPEI6, which represents a 6-month drought index. In this case, the ranger model continued to exhibit strong performance, with the highest  $R^2$  value of 0.512, and the lowest RSME (0.661) and MAE (0.522). The rf, svm, and cforest models closely followed. Once again, the CART and lars2 models were less accurate in predicting SPEI6 (Fig. A1c). For the 12-month time scale drought (SPEI12), the qrf model emerged as the top performer with the highest  $R^2$  value of 0.871, signifying its exceptional predictive accuracy of drought condition. The ranger, rf, and cforest models also demonstrated robust performance, while the CART and lars2 models exhibited comparatively lower accuracy in forecasting SPEI12 (Fig. A1d).

The Taylor diagram, a widely used pictorial tool, serves to evaluate the performance of ML models<sup>13,83</sup>. This diagram visualized the spatial pattern of calculated (reference field) against predicted (test field) multi-time scale SPEI values<sup>94</sup>. Fig. A2 depicts the Taylor's diagram, incorporating metrics such as RMSE, correlation coefficient, and standard deviation for SPEI1, SPEI3, SPEI6, and SPEI12. It was observed from the figure that all ML models exhibited a standard deviation of less than one across each SPEI time scale. Similarly, the correlation results varied from 0.40 to 0.80 for SPEI1, SPEI3, and SPEI6, while ranging from 0.40 to 0.95 for SPEI12. Additionally, centered RMSE was more scattered in SPEI12 compared to other time scales. Notably, the ranger model consistently outperformed than other models across all time scales.

### Observed and predicted performance of ML models for countrywide datasets

The scatter plot illustrating the performance of observed and predicted SPEIs is presented in Fig. A3a–d. For SPEI1 (Fig. A3a), actual vs predicted  $R^2$  values ranged from 0.26 to 0.69, with 46% (11 models) displaying  $R^2$  values equal to or exceeding 60%. Notably, models such as ranger, rf, qrf, and cforest demonstrated high accuracy during validation. In the case of SPEI3, 21% (5 models) exhibited an  $R^2$  value surpassing 0.60, with ranger, rf, qrf, cforest, and svm models showcasing exemplary prediction performance (Fig. A3b). Approximately 17% of SPEI6 models achieved an  $R^2$  value greater than 0.49, with ranger, svm, rf, and qrf models standing out (Fig. A3c). The  $R^2$  square values for SPEI12 ranged from 0.29 to 0.87, and during validation, qrf, ranger, and rf models exhibited  $R^2$  values exceeding 80%, indicating highly accurate prediction performance (Fig. A3d). In conclusion, the drought prediction performance of ML models varied across different time scales, with ranger, rf, and qrf models demonstrating consistency.

Another approach to identify the best model with fewer outlier issues is presented in Fig. A4a–d, providing an overview of the actual and predicted scenario of ML models, including outlier considerations. All models, across various time scales, showcased optimal performance when addressing the outlier problem. Empirical data supported the superiority of ranger, rf, and qrf models for precise drought prediction in Bangladesh. In contrast, CART, lars2, pcaNNet, ImStepAIC, and some other models struggled to overcome outlier values, negatively impacting their performance compared to the actual values of SPEI1, SPEI3, SPEI6, and SPEI12.

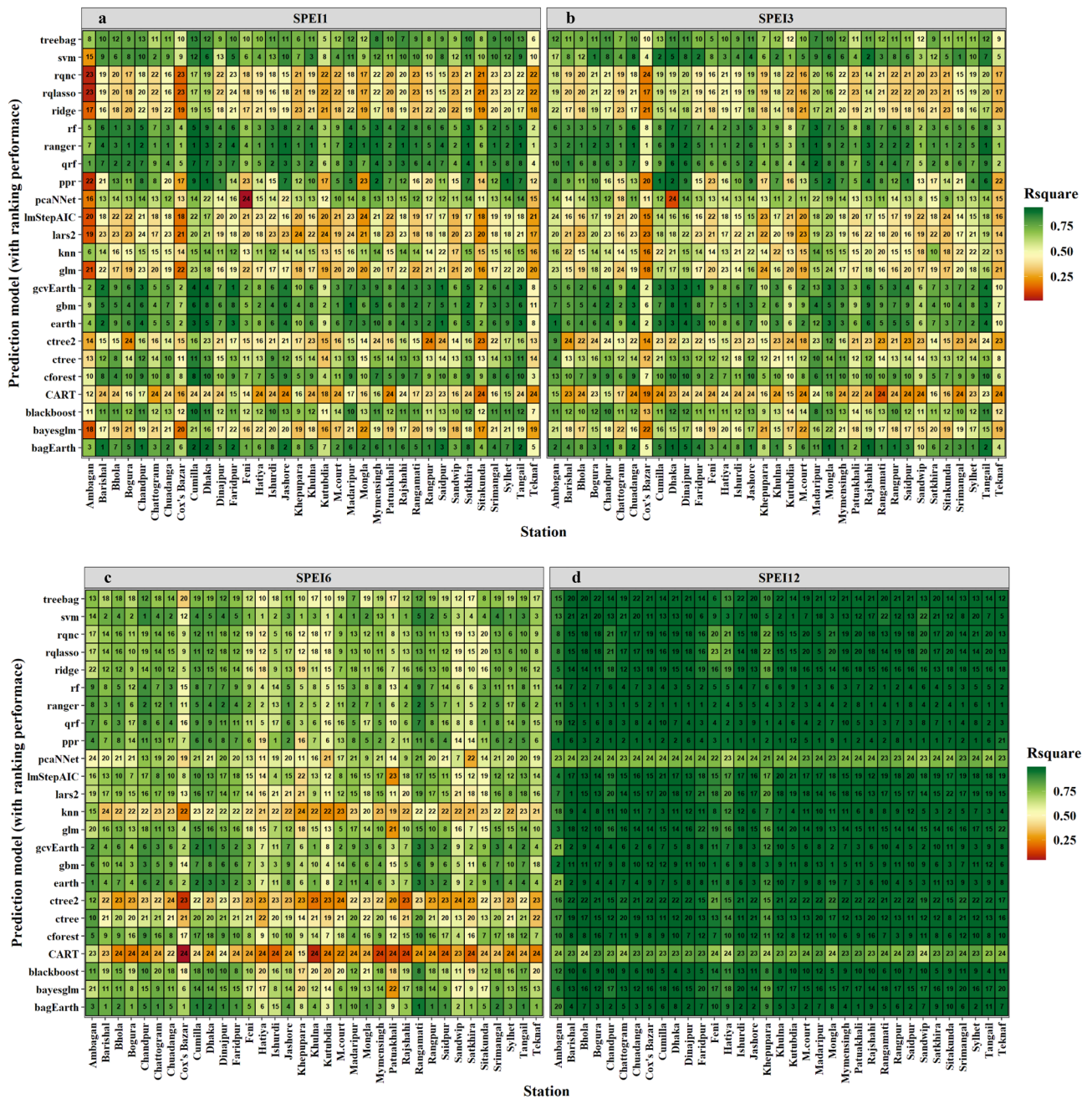
### Ranking the best predictive model for different regions at multiple timescales of SPEIs

Given the geographical positioning and climatic unpredictable changes across the country, a one-size-fits-all model would not be suitable for predicting drought in all locations. Here, we demonstrated a regional drought forecast for 35 meteorological stations using 24 ML models at various SPEI periods (Fig. 5a–d). The performance of each model was graded using higher  $R^2$  and lower MAE and RMSE values, illustrated through a heatmap. The results revealed that the best model differed across geographical locations and timespan. In the northern region of Bangladesh, for instance, at Rajshahi station in SPEI1 and SPEI12, the ranger model performed the best (ranked first). Similarly, the bagEarth and svm models had the highest performance in SPEI3 and SPEI6, respectively. In the southern region of Khulna, the ranger, bagEarth, earth, and ppr models performed most well at SPEI1, SPEI3, SPEI6, and SPEI12, respectively. Regarding regional representation, the ranger model demonstrated superior performance in 79% and 63% of regions (out of 35 stations) for SPEI1 and SPEI3, respectively. Conversely, for SPEI6 and SPEI12, the bagEarth and ppr models excelled, leading in 63% and 58% of regions, respectively (Fig. 6). However, the heatmaps of SPEI1 and SPEI3 were identical, and the usual  $R^2$  values for SPEI12 were fairly high. Refer to Figs. 5a–d and 6 for detailed information on the best model for each region.

### Evaluation of best predictive models for the specific region

We used scatter plots of fitted vs. observed values and  $R^2$  values to evaluate the best model for use across all regions and different time periods of SPEIs (Fig. 7). For SPEI1, the  $R^2$  values ranged from 0.57 to 0.93, indicating a positive correlation between the ML model and the observed data, with the model explaining 57–93 percent of the variance in the fitted data. Similarly, for SPEI3 and SPEI6, the  $R^2$  values ranged from 0.52 to 0.92 and 0.57 to 0.95 respectively, signifying a positive correlation between the ML models and the observed data, with the model's explaining 52–92 percent and 57–95 percent of the variance in the fitted data, respectively. Lastly, the high  $R^2$  value for SPEI12 suggested a better fit for the model. Thus, confirming the validity of the models selected for drought prediction in Bangladesh across different time scales and regions.

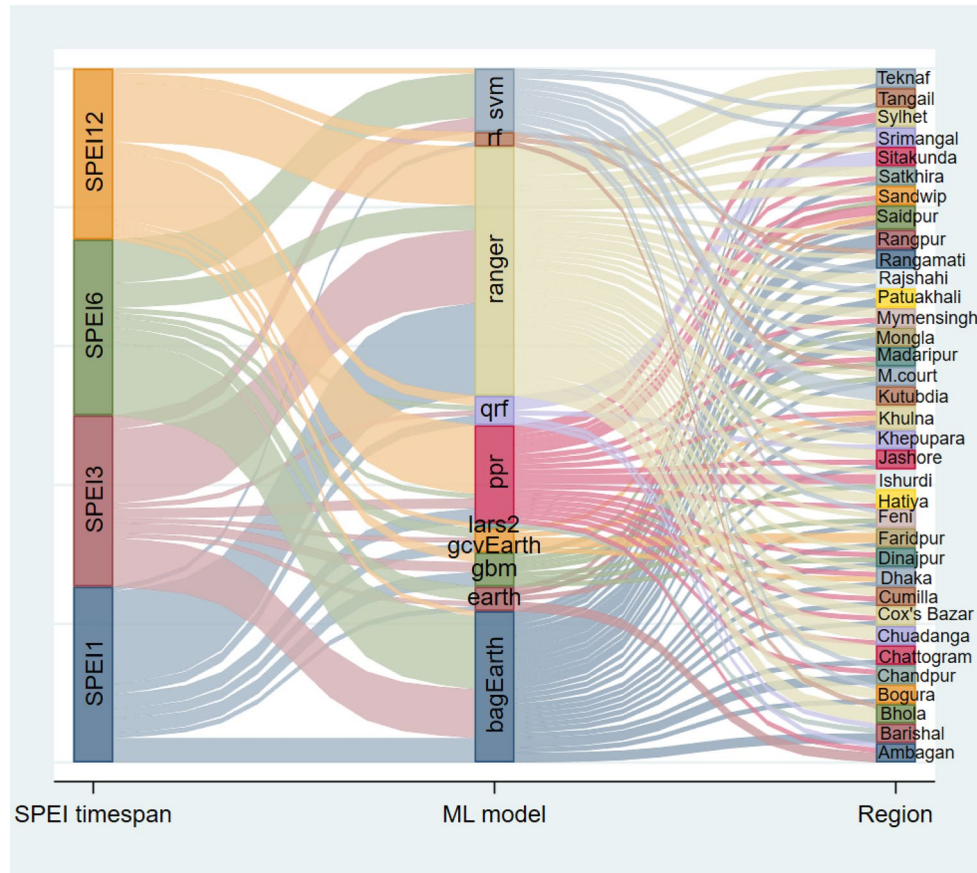
To examine the accuracy in SPEI prediction, boxplots of 25%, 50%, and 75% quantile values for both observed and projected SPEI are shown in Fig. 8. The figure illustrates that the identified best models adequately simulated the variability in SPEI1, SPEI3, SPEI6, and SPEI12 values across different regions. While many predicted SPEI values exhibited minimum fluctuation, except the observed values displayed a wide range when SPEI fell below -2 or exceeded 2 in a few cases. However, the identified best model showed better accuracy in simulating the variability and quantile of SPEIs compared to others. All prediction models exhibited enhanced performance in modeling SPEI quantiles across various SPEI scales, particularly at higher orders.



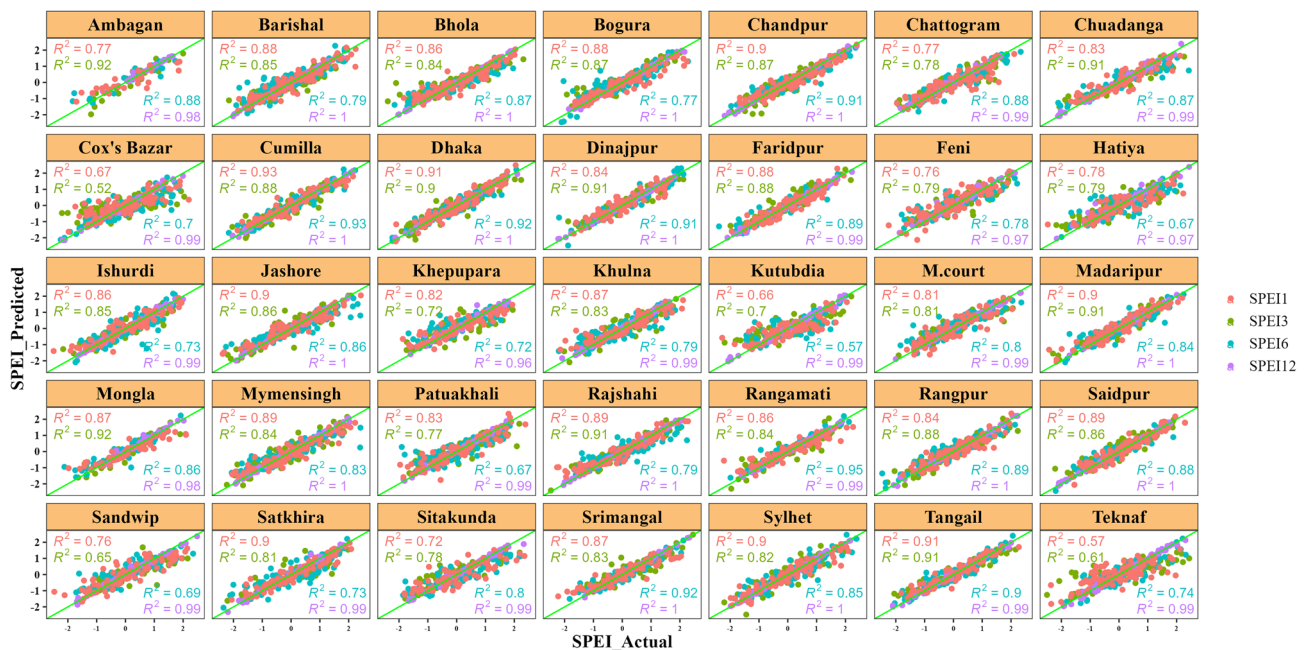
**Figure 5.** Heatmap illustrates the region-specific ML model selection for drought assessment based on  $R^2$ , MAE, and RMSE values. Various colors indicate the strength of the  $R^2$  values. The region-specific ranking of ML models for predicting (a) SPEI1, (b) SPEI3, (c) SPEI6, and (d) SPEI12 was displayed by the added value label in the middle of the box. Greater  $R^2$  and lower MAE and RMSE values defined the performance ranking scale of the model.

**Partial dependence plot for fitted projection**

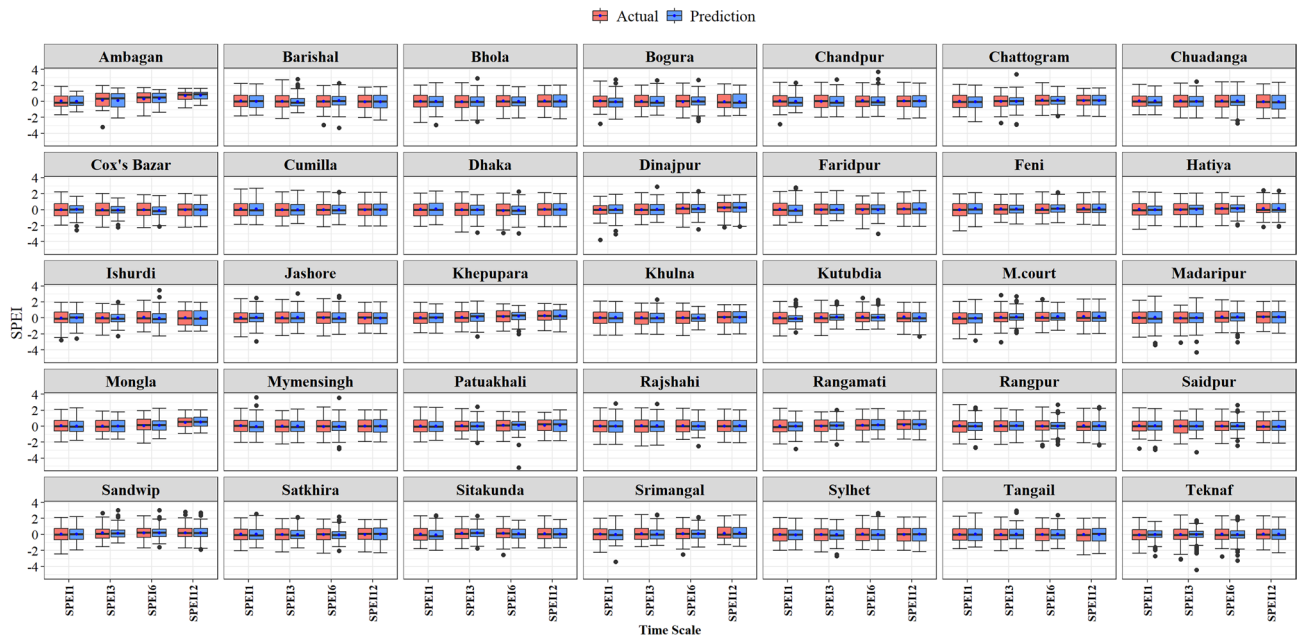
We employed a multivariate regression model to assess the significance of weather variables, and their relative effect on predicting SPEIs. In our analysis, ICE curves (depicted in black) and their mean (illustrated as the red line) were employed to visualize the relationships between individual weather attributes and the predicted SPEIs. This approach allowed us to identify critical climatic threshold values (Fig. 9). The findings revealed that Bangladesh experienced a range of drought moderate to severity levels, with a deficit of 92, 95, 115, and 143 mm of average rainfall over one, three, six, and twelve months, respectively. Temperature played a crucial role, with minimum, maximum, and mean temperatures exceeding  $20.7 \pm 1.1$ ,  $30.9 \pm 0.7$ , and  $25.9 \pm 0.8$  °C, respectively, resulting in severe drought conditions across these time scales. Similarly, we observed that extended periods of sunshine hours and relative humidity surpassing  $6.3 \pm 0.6$  h and  $77.3 \pm 1.3\%$ , respectively, contributed to drought conditions. Low wind speeds below  $1.9 \pm 0.2$  m/s and high evapotranspiration exceeding  $123 \pm 10$  mm at all



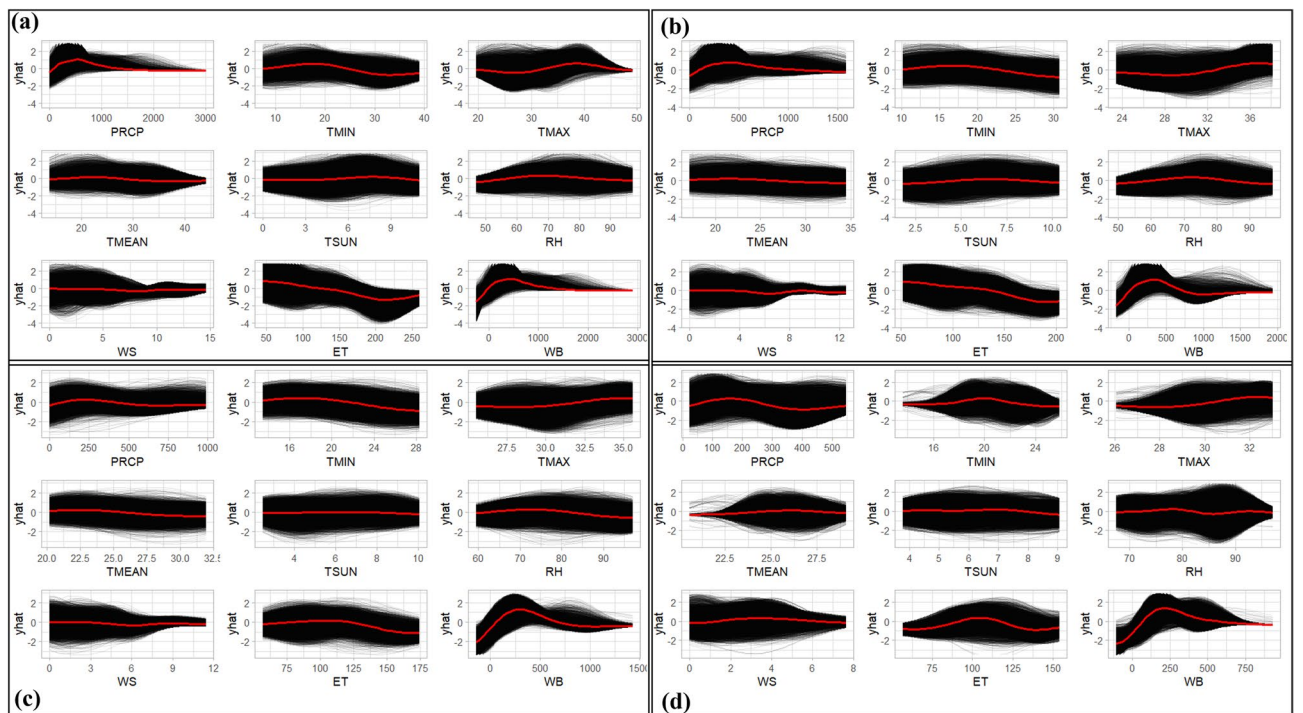
**Figure 6.** The Sankey graph illustrates a visualization of drought prediction models in Bangladesh across different timescales and regions.



**Figure 7.** Region-specific best predictive ML models' performance.



**Figure 8.** Box plot presentation of the best ML model performance of SPEI prediction for multiple time scales at the 35 investigated meteorological stations of Bangladesh.



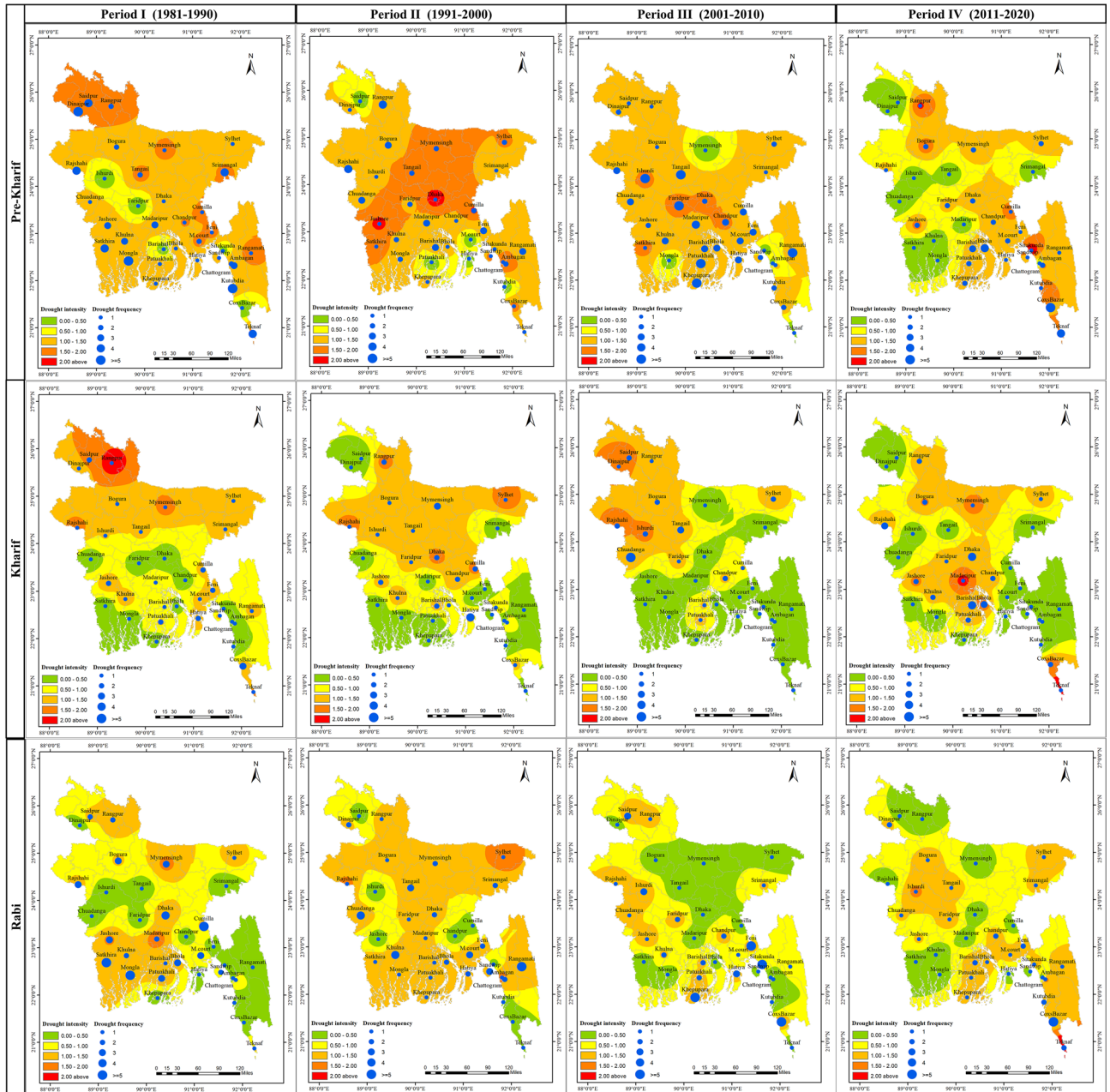
**Figure 9.** Fitted a partial dependence plot using the ICE curve method for each climatic feature against SPEI ( $\hat{y}$ ) for different time scales (a) SPEI1, (b) SPEI3, (c) SPEI6, and (d) SPEI12. Black and red curves denote ICE curves and their average value.

four-time scales also played a significant role in inducing drought in the country. Furthermore, the water balance was identified as a substantial factor affecting SPEI prediction. Below-average water balance levels, specifically 116, 143, 148, and 190 mm for one, three, six, and twelve months, respectively, were associated with drought occurrences in Bangladesh. These findings provide critical insights into the complex interplay of weather variables and their impact on drought patterns in the study region.

### Spatio-temporal pattern of seasonal drought intensity and frequency

Using the region-specific best selected model based on SPEI influential meteorological parameters, we predicted the seasonal intensity and frequency of drought over time in Bangladesh (Fig. 10). We divided the forty years into four periods, i.e., Period I: 1981–1990, Period II: 1991–2000, Period III: 2001–2010, and Period IV: 2011–2020. Results showed that while drought intensity has decreased over time, but the return period has become more frequent. Spatially, the drought intensity shifted from the northern to central and southern zones of the country. In periodic assessment, the period with the most severe drought intensity was Period II. Notably, the frequency of drought has increased in Periods III and IV, indicating an increase in the number of droughts that occurred twice a decade in the past.

Season-wise, Pre-kharif was the most common season for drought compared to other seasons. During the Period I of the Pre-kharif season, the northern, eastern, and a few southern regions of the country were primarily affected by severe drought. During Period II, the northern and the majority of the central regions were most



**Figure 10.** Spatio-temporal pattern of drought intensity and frequency based on the best predicted model for Pre-kharif, Kharif, and Rabi seasons over four decades (1981–1990, 1991–2000, 2001–2010, and 2011–2020) of Bangladesh. The authors used ArcGIS 10.8 (<https://www.arcgis.com/index.html>) to generate the map, employing the administrative shapefile of Bangladesh in the process. Shapefile republished from the Bangladesh Agricultural Research Council (BARC) database (<http://maps.barcapps.gov.bd/index.php>) under a CC BY license, with permission from Computer and GIS unit, BARC, original copyright 2014.

affected by drought, while the severity of drought in the north and center regions gradually relieved in Period III. In Period IV, the drought severity has been more prevalent in the Rangpur, Bogura, Sylhet, Mymensingh, Cumilla, Jashore, Sitakundu, Kutubdia and parts of southern regions in Bangladesh. Except for Rajshahi, the drought intensity of the Barind tract (located mainly in the northwestern part) was so unpredictable and has decreased significantly in recent decades. Noticeably, we found that the intensity of drought ( $> 1.0$ ) in Chattogram division has been affected continuously over the last forty years. The highest frequency of drought was observed during Period III (2001–2010). Among all periods, the spatial patterns of drought frequency had changed, and high crop-intensive areas had become more vulnerable in the Pre-kharif season.

During the Kharif season (June–October), the incidence of drought was lower than in the Pre-Kharif season, but the pattern was comparable. In Period I of the Kharif season, the northern region of the country was hit by a severe drought, and the central half of the region was affected by a moderate drought. The intensity of the drought shifted from the north to the central region during the succeeding decade. During Period III, the majority of the northwest region again witnessed a severe drought, while the rest of the county was affected by a mild drought. In Period IV, the frequency of drought increased relative to previous periods, and its intensity rose in the north-eastern, central, and southern regions of Bangladesh.

During the Rabi season in Period I, the southwest and a portion of the northern region experienced drought intensity larger than one, and drought frequency greater than three times the average. However, nearly the entire country faced drought conditions in the succeeding decade. Throughout the country, the drought intensity has reduced, but frequency increased in Period III. Again, during Period IV, the divisions of Barishal, Chattogram, and Sylhet, as well as parts of the central regions, had droughts with intensities more than one. This suggests that drought conditions were similarly erratic throughout the Rabi season.

Overall, the intensity and frequency of drought in Bangladesh have exhibited erratic patterns over the past forty years, with certain periods witnessing more severe droughts than others. Factors such as average rainfall, temperature, sunshine hours, relative humidity, wind speed, and evapotranspiration have collectively influenced drought intensity, along with water balance deficits. These factors have significantly influenced SPEI predictions, rendering regions with high crop intensity more susceptible. This underscores the crucial importance of comprehending the impacts of drought on food production and livelihoods in the region.

## Discussion

Bangladesh experiences a predominantly tropical climate characterized by high temperatures and humidity, with droughts being more prevalent than other climate stressors in the country<sup>13</sup>. The occurrence, severity, and duration of droughts vary based on meteorological, hydrological, and agricultural factors<sup>95</sup>. In an effort to develop best ML-based models for predicting seasonal droughts in Bangladesh, this study utilized data from 35 meteorological stations at four distinct time scales. Applying 24 ML models based on the SPEI, we compared them with individual station predictions at different time intervals, contributing to the formulation of a spatio-temporal drought management program. In alignment with previous studies by Alamgir et al.<sup>96</sup> and Yaseen et al.<sup>13</sup>, our SPEI1 results indicated a short-term rainfall deficit, SPEI3 and SPEI6 suggested agricultural drought, and SPEI12 reflected reductions in river flow and groundwater levels, signifying hydrological droughts in Bangladesh.

The use of ML models guaranteed the robustness of the drought prediction. We observed precipitation, maximum and minimum temperature, and water balance was the essential elements for predicting droughts in Bangladesh. Similar findings have been found by Rahman and Lateh<sup>8</sup>. In Pakistan, relative humidity, temperature, and wind speed were the most significant meteorological characteristics for accurately predicting drought<sup>97</sup>. Using ML models, Zhang et al.<sup>46</sup> determined that temperature and precipitation, air pressure, wind speed, relative humidity, and duration of sunshine have a substantial effect on drought in China. So, it is evident that the relative importance of meteorological parameters on drought prediction varies among geo-climatic and geographical segments. Zhang et al.<sup>46</sup> also stated that rf, ranger, qrf, bagEarth, and svm were the best methods for annual drought prediction. In the context of model stability assessment, the ranger model has the highest chance for 1–5 rating 90% of the time in SPEI1, SPEI3, and SPEI12 time scales. On another side, svm and bagEarth have 91% and 86% probability on the SPEI6 time scale, respectively. The svm-based models exhibited great temporal and geographical drought characterization in Pakistan<sup>97</sup>. Similarly, we observed that the performance and efficacy of the model varied based on the time frame and location of its application.

Our exemplary findings urged for the adoption of regional models instead of a single unified model, considering regional spatio-temporal heterogeneity. This conclusion is drawn from the analysis of influential meteorological variables on the SPEI depicted in Fig. 4, and the identification of region-specific best models shown in Fig. 6. These models indicate a spatio-temporal shift in both the intensity and frequency of seasonal drought patterns in Bangladesh, as illustrated in Fig. 10. Notably, the intensity of drought in climate hotspot regions, such as northwest and northern Bangladesh, has diminished and gradually shifted towards the center and south. Similar findings were reported by Mohsenipour et al.<sup>12</sup>. Projections until the end of the century (the 2070s) anticipate a decrease in maximum drought intensity<sup>98</sup>. Among the three seasons in Bangladesh, the Pre-kharif season witnessed the most significant decline in drought severity over the years. Comparable changes in drought patterns have been reported periodically in several Asian countries, including Nepal, Bhutan, Cambodia, Lao PDR, India, China, and Pakistan<sup>1,46,97,99–101</sup>.

Considering land use changes in drought prediction models is essential for a comprehensive understanding of drought dynamics. It enables improved assessments of water availability, vegetation responses, human-induced vulnerabilities, feedback mechanisms, and the design of effective adaptation and mitigation strategies<sup>102,103</sup>. By accounting for land use changes, enhance the accuracy and relevance of drought predictions, ultimately supporting sustainable water resource management and resilience to drought events.

Changes in drought duration and recurrent nature justified the trends in crop damage and lower cropping intensity in Bangladesh<sup>8</sup>. The Climate Change Cell of Bangladesh<sup>104</sup> reported that droughts reduced T. Aman rice yields by 45–60% and Rabi crop yields by 50–70%. Boro (grown in the Rabi season), the major rice, is entirely dependent on irrigation. Prolonged water scarcity encourages excessive groundwater extraction for irrigation, which further depletes groundwater levels in drought-prone regions<sup>9</sup>. To cope with drought, many farmers have chosen drought-resistant rice varieties cultivation<sup>105,106</sup>, but adoption rates remain low<sup>107</sup>. Other adaptation strategies included switching farming practices or changing the crop-sowing windows. Depending on the severity of the drought, some rural households may choose to migrate or change their livelihoods<sup>108</sup>. Overall, this could severely impact food supplies and endanger food security.

The economic losses associated with severe drought in specific countries can vary widely depending on factors such as the severity of the drought, the country's economic resilience, and its dependence on agriculture and water resources<sup>109</sup>. The United States experienced substantial economic losses from drought, particularly in agricultural states such as California and Texas. Prolonged droughts can lead to reduced crop yields, increased irrigation costs, and even water shortages for urban areas. The economic impact can run into billions of dollars<sup>110</sup>. Australia has a history of severe droughts, which can devastate the agriculture sector. The "Millennium Drought" in the early 2000s, for example, resulted in significant economic losses, affecting everything from livestock farming to wine production<sup>111</sup>. Severe droughts in South Africa have led to water scarcity and reduced agricultural productivity. In recent years, the country has experienced drought-related losses in key sectors like maize production and livestock farming<sup>112</sup>. India, with its large agricultural sector, is highly vulnerable to drought. Severe droughts can lead to crop failures, food shortages, and economic hardships for farmers. The economic losses can be substantial<sup>113</sup>. Droughts in Brazil impacted its important agricultural and livestock sectors. The country's economy is closely tied to these industries, making drought-related losses a significant concern<sup>114</sup>. In countries with fragile economies like Somalia, severe drought had devastating effects, leading to food shortages, loss of livestock, and economic distress<sup>115</sup>. Ethiopia experienced recurring droughts, leading to food insecurity and economic challenges. Efforts to mitigate the impact of drought and build resilience are ongoing in the country<sup>116</sup>. In sum, the imperative to prioritize drought mitigation cannot be overstated for nations grappling with the profound impacts of water scarcity. Making drought a top policy concern is paramount to developing comprehensive strategies encompassing preparedness, response, and recovery. As evident in the experiences of these countries, a forward-looking policy framework is crucial to building resilience, ensuring effective response mechanisms, and facilitating a swift recovery from the complexities posed by drought. By placing drought mitigation at the forefront of policy agendas, nations can proactively address the multifaceted challenges posed by water scarcity, safeguarding their economies, food security, and overall well-being for a more sustainable and resilient future.

The potential for future research on long-term drought prediction is vast and holds great significance in understanding and mitigating the impacts of drought events. In this regard, deep learning models, specifically transformers, could be used on drought prediction. Developing and refining climate models to better capture long-term climate variability and teleconnections can enhance our ability to predict droughts beyond seasonal time scales<sup>117</sup>. Incorporating more accurate representations of physical processes, feedback mechanisms, and regional climate dynamics can improve model performance for longer-term predictions<sup>118</sup>. Incorporating data from remote sensing, ground-based observations, and hydrological models can enhance our understanding of the complex feedbacks between the land surface and the atmosphere, enabling better predictions of long-term drought conditions<sup>119</sup>. Ensemble methods, data assimilation techniques, and hybrid models that combine the strengths of physical models and data-driven approaches may improve the skill and reliability of long-term drought predictions<sup>120,121</sup>. So, the understanding of long-term drought prediction and contribute to more effective planning, preparedness, and management of drought events at longer time scales is required.

## Conclusions and policy recommendations

The objective of this research is to determine the most effective machine learning methods and categorize the key factors influencing drought prediction. Utilizing a dataset spanning four decades of different weather variables, we conducted extensive statistical analysis and employed data visualization techniques to gain insights from the data. We explored a total of twenty-four machine learning models, encompassing both regression and classification approaches, which included linear, non-linear, and ensemble learning techniques. An inclusive assessment of the twenty-four models' performance, by applying statistical metrics such as the  $R^2$ , RMSE, and MAE method during the model validation phase. Results shows, four best ML methods, ranger, bagEarth, support vector machine, and random forest have been identified for the prediction of multi-scale drought indices. Our findings revealed that Bangladesh experienced varying levels of drought, ranging from moderate to severe and existing the shifting tendency by regions with specific deficits in average rainfall over different time periods. For instance, the deficit was 92, 95, 115, and 143 mm over 1, 3, 6, and 12-months' time span, respectively. Temperature was identified as a crucial factor influencing drought conditions, with minimum, maximum, and mean temperatures exceeding certain thresholds resulting in severe drought conditions across different time spans. This information is valuable for understanding how temperature variations can impact drought severity. Additionally, we observed that extended periods of sunshine hours and high relative humidity levels also contributed to drought conditions in the region. Low wind speeds and high evapotranspiration further exacerbated drought conditions.

Ranking of multi-model machine learning algorithms indicate the appropriate selective options for different stations getting best prediction accuracy during drought forecasting. Reliable and effective strategies for choosing the right weather variables and constructing accurate predictive models can help to reduce the destructive impacts of drought. In this study, the application of ML models confirmed the efficacy and reliability of the research. Moreover, model selection for each region improves the study performance, and the output can be used to assess drought risk in Bangladesh. Our research proposes specific recommendations to address various



aspects of drought mitigation in Bangladesh. To enhance water resource management, we advocate for strategies that improve water storage, allocation, and distribution systems, including the construction of reservoirs, promotion of water recycling and rainwater harvesting, and implementation of water conservation practices. In terms of promoting agricultural resilience, we highlight the importance of adopting climate-smart agricultural practices such as cultivating drought-tolerant crop varieties, implementing efficient irrigation techniques, practicing crop rotation, promoting agroforestry, and adopting sustainable land management practices. Additionally, we emphasize the significance of strengthening early warning systems by integrating climate data, remote sensing technologies, and advanced modeling techniques to improve the accuracy and lead time of drought predictions. Furthermore, our recommendations aim to enhance drought preparedness and response through the development of drought contingency plans, establishment of drought monitoring and assessment frameworks, and provision of financial and technical support to vulnerable communities. Lastly, we advocate for the formulation and implementation of policies that prioritize drought risk reduction and sustainable water resource management, including the integration of drought mitigation strategies into national and regional development plans, establishment of regulatory frameworks, and allocation of financial resources for drought resilience projects. By including these specific recommendations, our study aims to bridge the gap between research findings and practical applications, providing policymakers and stakeholders with tangible guidance to effectively mitigate the impacts of droughts in Bangladesh.

Future research should seek to better understand the effects of climate change on drought intensity, impact mechanism on crop, and explore strategies for adapting to longer-term drought. Collaboration between researchers, local authorities, and community organizations can help to identify effective solutions that take into account local socio-economic contexts and water management strategies that are amenable to research. Additionally, more research is needed to assess mitigation strategies, such as better water infrastructure and conservation efforts, to lessen the impacts of droughts.

### Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

### Code availability

The study primarily used the following R packages: "caret" and "ggplot2" of programming R version 4.2.1.

Received: 22 August 2023; Accepted: 30 December 2023

Published online: 04 January 2024

### References

- Miyan, M. A. Droughts in Asian least developed countries: Vulnerability and sustainability. *Weather Clim. Extrem.* **7**, 11 (2015).
- Campos, J. N. B. Paradigms and public policies on drought in northeast Brazil: A historical perspective. *Environ. Manag.* **55**, 1052–1063 (2015).
- Hao, Z. & Singh, V. P. Drought characterization from a multivariate perspective: A review. *J. Hydrol.* **527**, 668–678 (2015).
- Orimoloye, I. R., Belle, J. A., Orimoloye, Y. M., Olusola, A. O. & Olojede, O. O. Drought: A common environmental disaster. *Atmosphere (Basel)* **13**, 111 (2022).
- Thomas, J. & Prasannakumar, V. Temporal analysis of rainfall (1871–2012) and drought characteristics over a tropical monsoon-dominated State (Kerala) of India. *J. Hydrol.* **534**, 266–280 (2016).
- Mare, F., Bahta, Y. T. & Van Niekerk, W. The impact of drought on commercial livestock farmers in South Africa. *Dev. Pract.* **28**, 884–898 (2018).
- Afrin, S. *Livelihood Pattern of Flooded Families in Bangladesh: A Study in Two Villages*. 1–131 (2015).
- Rahman, M. R. & Lateh, H. Meteorological drought in Bangladesh: Assessing, analysing and hazard mapping using SPI, GIS and monthly rainfall data. *Environ. Earth Sci.* **75**, 22 (2016).
- Habiba, U., Shaw, R. & Takeuchi, Y. Drought risk reduction through a socio-economic, institutional and physical approach in the northwestern region of Bangladesh. *Environ. Hazards* **10**, 31 (2011).
- Mortuza, M. R., Moges, E., Demissie, Y. & Li, H.-Y. Historical and future drought in Bangladesh using copula-based bivariate regional frequency analysis. *Theor. Appl. Climatol.* **135**, 855–871 (2019).
- Mishra, A. K. & Singh, V. P. Drought modeling—A review. *J. Hydrol.* **403**, 157–175 (2011).
- Mohsenipour, M., Shahid, S., Chung, E. & Wang, X. Changing pattern of droughts during cropping seasons of Bangladesh. *Water Resour. Manag.* **32**, 3 (2018).
- Yaseen, Z. M., Ali, M., Sharafati, A., Al-Ansari, N. & Shahid, S. Forecasting standardized precipitation index using data intelligence models: Regional investigation of Bangladesh. *Sci. Rep.* **11**, 1–25 (2021).
- Bhaga, T. D., Dube, T., Shekede, M. D. & Shoko, C. Impacts of climate variability and drought on surface water resources in sub-Saharan Africa using remote sensing: A review. *Remote Sens.* **12**, 33 (2020).
- Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscale drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Clim.* **23**, 1696–1718 (2010).
- Mondol, M. A. H., Ara, I. & Das, S. C. Meteorological drought index mapping in Bangladesh using standardized precipitation index during 1981–2010. *Adv. Meteorol.* **2017**, 4642060 (2017).
- Al-Mamun, A. *et al.* Identification of meteorological drought prone area in Bangladesh using standardized precipitation index. *J. Earth Sci. Clim. Change* **9**, 1000457 (2018).
- Moazzam, M. F. U., Rahman, G., Munawar, S., Farid, N. & Lee, B. G. Spatiotemporal rainfall variability and drought assessment during past five decades in South Korea using SPI and SPEI. *Atmosphere (Basel)* **13**, 292 (2022).
- Liu, C., Yang, C., Yang, Q. & Wang, J. Spatiotemporal drought analysis by the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI) in Sichuan Province, China. *Sci. Rep.* **11**, 1–14 (2021).
- Ojha, S. S., Singh, V. & Roshni, T. Comparison of meteorological drought using SPI and SPEI. *Civ. Eng. J.* **7**, 2130–2149 (2021).
- Wang, H., Chen, Y., Pan, Y., Chen, Z. & Ren, Z. Assessment of candidate distributions for SPI/SPEI and sensitivity of drought to climatic variables in China. *Int. J. Climatol.* **39**, 4392–4412 (2019).
- McKee, T. B. Drought monitoring with multiple time scales. In *Proceedings of 9th Conference on Applied Climatology*, Boston, 1995 (1995).

23. Botai, C. M., Botai, J. O., Dlamini, L. C., Zwane, N. S. & Phaduli, E. Characteristics of droughts in South Africa: A case study of free state and north west provinces. *Water* **8**, 133 (2016).
24. Temoso, O., Villano, R. A. & Hadley, D. Agricultural productivity, efficiency and growth in a semi-arid country: A case study of Botswana, African. *J. Agric. Resour. Econ.* **10**, 192–206 (2015).
25. Bhunia, P., Das, P. & Maiti, R. Meteorological drought study through SPI in three drought prone districts of West Bengal, India. *Earth Syst. Environ.* **4**, 43–55 (2020).
26. Naresh Kumar, M., Murthy, C. S., Sesa Sai, M. V. R. & Roy, P. S. On the use of standardized precipitation index (SPI) for drought intensity assessment. *Meteorol. Appl. J. Forecast. Pract. Appl. Train. Tech. Model.* **16**, 381–389 (2009).
27. Danandeh Mehr, A., Sorman, A. U., Kahya, E. & Hesami Afshar, M. Climate change impacts on meteorological drought using SPI and SPEI: Case study of Ankara, Turkey. *Hydrol. Sci. J.* **65**, 254–268 (2020).
28. Li, L., She, D., Zheng, H., Lin, P. & Yang, Z.-L. Elucidating diverse drought characteristics from two meteorological drought indices (SPI and SPEI) in China. *J. Hydrometeorol.* **21**, 1513–1530 (2020).
29. Lee, S.-H., Yoo, S.-H., Choi, J.-Y. & Bae, S. Assessment of the impact of climate change on drought characteristics in the Hwanghae Plain, North Korea using time series SPI and SPEI: 1981–2100. *Water* **9**, 579 (2017).
30. Chen, S., Zhang, L., Liu, X., Guo, M. & She, D. The use of SPEI and TVDI to assess temporal-spatial variations in drought conditions in the middle and lower reaches of the Yangtze River Basin, China. *Adv. Meteorol.* **2018**, 9362041 (2018).
31. Bae, S., Lee, S.-H., Yoo, S.-H. & Kim, T. Analysis of drought intensity and trends using the modified SPEI in South Korea from 1981 to 2010. *Water* **10**, 327 (2018).
32. Jia, Y., Zhang, B. & Ma, B. Daily SPEI reveals long-term change in drought characteristics in Southwest China. *Chinese Geogr. Sci.* **28**, 680–693 (2018).
33. Yang, M., Yan, D., Yu, Y. & Yang, Z. SPEI-based spatiotemporal analysis of drought in Haihe River Basin from 1961 to 2010. *Adv. Meteorol.* **2016**, 7658015 (2016).
34. Ye, L. *et al.* Spatio-temporal analysis of drought indicated by SPEI over northeastern China. *Water* **11**, 908 (2019).
35. Musei, S. K., Nyaga, J. M. & Dubow, A. Z. SPEI-based spatial and temporal evaluation of drought in Somalia. *J. Arid Environ.* **184**, 104296 (2021).
36. Mutsotso, R. B., Sichangi, A. W. & Makokha, G. O. *Spatio-Temporal Drought Characterization in Kenya from 1987 to 2016* (2018).
37. Alam, N. M. *et al.* Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India. *Phys. Chem. Earth Parts A/B/C* **100**, 31–43 (2017).
38. Bera, B., Shit, P. K., Sengupta, N., Saha, S. & Bhattacharjee, S. Trends and variability of drought in the extended part of Chhota Nagpur plateau (Singbhum Protocontinent), India applying SPI and SPEI indices. *Environ. Challenges* **5**, 100310 (2021).
39. Dai, A. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *J. Geophys. Res. Atmos.* **116**, 21 (2011).
40. Karl, T. R. The sensitivity of the Palmer Drought Severity Index and Palmer's Z-index to their calibration coefficients including potential evapotranspiration. *J. Clim. Appl. Meteorol.* **2**, 77–86 (1986).
41. Tan, C., Yang, J. & Li, M. Temporal-spatial variation of drought indicated by SPI and SPEI in Ningxia Hui Autonomous Region, China. *Atmosphere (Basel)* **6**, 1399–1421 (2015).
42. Potop, V., Možný, M. & Soukup, J. Drought evolution at various time scales in the lowland regions and their impact on vegetable crops in the Czech Republic. *Agric. For. Meteorol.* **156**, 121–133 (2012).
43. Mokhtar, A. *et al.* Estimation of SPEI meteorological drought using machine learning algorithms. *IEEE Access* **9**, 65503–65523 (2021).
44. Abdel-Fattah, M. K., Mokhtar, A. & Abdo, A. I. Application of neural network and time series modeling to study the suitability of drain water quality for irrigation: A case study from Egypt. *Environ. Sci. Pollut. Res.* **28**, 898–914 (2021).
45. Mossad, A. & Alazba, A. A. Drought forecasting using stochastic models in a hyper-arid climate. *Atmosphere (Basel)* **6**, 410–430 (2015).
46. Zhang, R., Chen, Z. Y., Xu, L. J. & Ou, C. Q. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. *Sci. Total Environ.* **665**, 31 (2019).
47. Elbeltagi, A. *et al.* The impact of climate changes on the water footprint of wheat and maize production in the Nile Delta, Egypt. *Sci. Total Environ.* **743**, 140770 (2020).
48. Maroufpoor, S., Maroufpoor, E., Bozorg-Haddad, O., Shiri, J. & Mundher Yaseen, Z. Soil moisture simulation using hybrid artificial intelligent model: Hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm. *J. Hydrol.* **575**, 544–556 (2019).
49. Ganguli, P. & Reddy, M. J. Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrol. Process.* **28**, 4989–5009 (2014).
50. Feng, Y., Cui, N., Chen, Y., Gong, D. & Hu, X. Development of data-driven models for prediction of daily global horizontal irradiance in northwest China. *J. Clean. Prod.* **223**, 136–146 (2019).
51. Granata, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric. Water Manag.* **217**, 303–315 (2019).
52. Mouatadid, S., Raj, N., Deo, R. C. & Adamowski, J. F. Input selection and data-driven model performance optimization to predict the Standardized Precipitation and Evaporation Index in a drought-prone region. *Atmos. Res.* **212**, 130–149 (2018).
53. Masinde, M. Artificial neural networks models for predicting effective drought index: Factoring effects of rainfall variability. *Mitig. Adapt. Strategy Glob. Chang.* **19**, 1139–1162 (2014).
54. Belayneh, A., Adamowski, J., Khalil, B. & Ozga-Zielinski, B. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J. Hydrol.* **508**, 418–429 (2014).
55. Djerbouai, S. & Souag-Gamane, D. Drought forecasting using neural networks, wavelet neural networks, and stochastic models: Case of the Algerois Basin in North Algeria. *Water Resour. Manag.* **30**, 2445–2464 (2016).
56. Deo, R. C. & Şahin, M. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos. Res.* **3**, 161–162 (2015).
57. Anshuka, A., van Ogtrop, F. F. & Willem Vervoort, R. Drought forecasting through statistical models using standardised precipitation index: A systematic review and meta-regression analysis. *Nat. Hazards* **97**, 65 (2019).
58. Ahmed, K., Shahid, S., Chung, E.-S., Wang, X. & Harun, S. B. Climate change uncertainties in seasonal drought severity-area-frequency curves: Case of arid region of Pakistan. *J. Hydrol.* **570**, 473–485 (2019).
59. Elbeltagi, A. *et al.* Estimating the standardized precipitation evapotranspiration index using data-driven techniques: A regional study of Bangladesh. *Water (Switzerland)* **14**, 1–16 (2022).
60. Kafy, A. *et al.* Assessment and prediction of index based agricultural drought vulnerability using machine learning algorithms. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2023.161394> (2023).
61. Osmani, S. A. *et al.* Prediction of monthly dry days with machine learning algorithms: A case study in Northern Bangladesh. *Sci. Rep.* **12**, 1–21 (2022).
62. Rahman, M. R. & Lateh, H. Climate change in Bangladesh: A spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model. *Theor. Appl. Climatol.* **128**, 27–41 (2017).
63. Shahid, S. Recent trends in the climate of Bangladesh. *Clim. Res.* **42**, 185–193 (2010).

64. Pereira, J. J., Pulhin, J. M. & Shaw, R. *Climate Change Adaptation and Disaster Risk Reduction: An Asian Perspective* (Emerald Group Pub. Limited, 2010).
65. Dastagir, M. R. Modeling recent climate change induced extreme events in Bangladesh: A review. *Weather Clim. Extrem.* **7**, 49–60 (2015).
66. Abedin, M. A., Collins, A. E., Habiba, U. & Shaw, R. Climate change, water scarcity, and health adaptation in southwestern coastal Bangladesh. *Int. J. Disaster Risk Sci.* **10**, 28–42 (2019).
67. Pande, C. B., Al-ansari, N., Kushwaha, N. L., Srivastava, A. & Noor, R. *Forecasting of SPI and Meteorological Drought Based on the Artificial Neural Network and MSP Model Tree*. 1–24 (2022).
68. Buttafuoco, G., Caloiero, T. & Coscarelli, R. Analyses of drought events in Calabria (southern Italy) using standardized precipitation index. *Water Resour. Manag.* **29**, 557–573 (2015).
69. Beucler, T., Ebert-Uphoff, I., Rasp, S., Pritchard, M. & Gentine, P. *Machine Learning for Clouds and Climate* (Clouds Climate Cambridge University Press Cambridge, 2020).
70. Mullick, M. R. A., Nur, M. R. M., Alam, M. J. & Islam, K. M. A. Observed trends in temperature and rainfall in Bangladesh using pre-whitening approach. *Glob. Planet. Change* <https://doi.org/10.1016/j.gloplacha.2018.10.001> (2019).
71. Chiew, F. H. S., Kamaladasa, N. N., Malano, H. M. & McMahon, T. A. Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agric. Water Manag.* **28**, 9–21 (1995).
72. Mokhtar, A. *et al.* Evapotranspiration as a response to climate variability and ecosystem changes in southwest, China. *Environ. Earth Sci.* **79**, 312 (2020).
73. Li, Y., Ren, F., Li, Y., Wang, P. & Yan, H. Characteristics of the regional meteorological drought events in Southwest China during 1960–2010. *J. Meteorol. Res.* **28**, 381–392 (2014).
74. Jiang, W., Wang, L., Feng, L., Zhang, M. & Yao, R. Drought characteristics and its impact on changes in surface vegetation from 1981 to 2015 in the Yangtze River Basin, China. *Int. J. Climatol.* **40**, 3380–3397 (2020).
75. Nafarzadegan, A. R. *et al.* Drought area monitoring during the past three decades in Fars province, Iran. *Quat. Int.* **250**, 27–36 (2012).
76. Banimahd, S. A. & Khalili, D. Factors influencing Markov chains predictability characteristics, utilizing SPI, RDI, EDI and SPEI drought indices in different climatic zones. *Water Resour. Manag.* **27**, 3911–3928 (2013).
77. Spinoni, J., Naumann, G., Carrao, H., Barbosa, P. & Vogt, J. World drought frequency, duration, and severity for 1951–2010. *Int. J. Climatol.* **34**, 2792–2804 (2014).
78. Bouktif, S., Fiaz, A., Ouni, A. & Serhani, M. A. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* **11**, 22 (2018).
79. Khan, M. J. U. *et al.* Observed trends in climate extremes over Bangladesh from 1981 to 2010. *Clim. Res.* **77**, 45–61 (2019).
80. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
81. Martinez-Taboada, F. & Redondo, J. I. Variable importance plot (mean decrease accuracy and mean decrease Gini). *PLoS One* **15**, e0230799 (2020).
82. Han, H., Guo, X. & Yu, H. Variable selection using mean decrease accuracy and mean decrease Gini based on random forest. *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS* **7**, 219–224 (2016).
83. Singh, U. K., Jamei, M., Karbasi, M., Malik, A. & Pandey, M. Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. *J. Hydrol.* **607**, 127549 (2022).
84. Janizadeh, S. *et al.* Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. *J. Environ. Manag.* **298**, 113551 (2021).
85. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning*. Vol. 112 (Springer, 2013).
86. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
87. Ebrahimi-Khusfi, Z., Nafarzadegan, A. R. & Dargahian, F. Predicting the number of dusty days around the desert wetlands in southeastern Iran using feature selection and machine learning techniques. *Ecol. Indic.* **125**, 107499 (2021).
88. Kuhn, M. *et al.* Classification and regression tree. *Environ. Ecol. Stat. R* <https://doi.org/10.1201/b17172-15> (2021).
89. Chen, Y. *et al.* Toward the development of deep learning analyses for snow avalanche releases in mountain regions. *Geocarto Int.* **3**, 1–26 (2021).
90. Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H. R. & Van Oost, K. Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci. Total Environ.* **664**, 1117–1132 (2019).
91. Nabavi-Pelesaraei, A., Rafiee, S., Mohtasebi, S. S., Hosseinzadeh-Bandbafha, H. & Chau, K. Integration of artificial intelligence methods and life cycle assessment to predict energy output and environmental impacts of paddy production. *Sci. Total Environ.* **631–632**, 1279–1294 (2018).
92. Greenwell, B. M. pdp: An R package for constructing partial dependence plots. *R J.* **9**, 421–436 (2017).
93. Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat.* **24**, 44–65 (2015).
94. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **106**, 7183–7192 (2001).
95. Heim, R. R. A review of twentieth-century drought indices used in the United States. *Bull. Am. Meteorol. Soc.* **83**, 31 (2002).
96. Alamgir, M. *et al.* Parametric assessment of seasonal drought risk to crop production in Bangladesh. *Sustainable* **11**, 1–17 (2019).
97. Khan, N. *et al.* Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.* **139**, 103562 (2020).
98. Kamruzzaman, M., Jang, M. W., Cho, J. & Hwang, S. Future changes in precipitation and drought characteristics over Bangladesh under CMIP5 climatological projections. *Water (Switzerland)* **11**, 291 (2019).
99. Wang, Z. *et al.* Does drought in China show a significant decreasing trend from 1961 to 2009?. *Sci. Total Environ.* **579**, 4 (2017).
100. Mishra, A. & Liu, S. C. Changes in precipitation pattern and risk of drought over India in the context of global warming. *J. Geophys. Res.* **119**, 13 (2014).
101. Ge, Y., Cai, X., Zhu, T. & Ringler, C. Drought frequency change: An assessment in northern India plains. *Agric. Water Manag.* **176**, 111 (2016).
102. Afzal, M. & Ragab, R. Drought risk under climate and land use changes: Implication to water resource availability at catchment scale. *Water* **11**, 1–29 (2019).
103. Sajikumar, N. & Sobhana, R. Impact of land cover and land use change on runoff characteristics. *J. Environ. Manag.* **161**, 460 (2015).
104. MoEF. *Bangladesh Climate Change Strategy and Action Plan 2009*. <http://nda.erd.gov.bd/files/1/Publications/CCPolicyDocuments/BCCSAP2009.pdf> (2009).
105. Arora, A., Bansal, S. & Ward, P. S. Do farmers value rice varieties tolerant to droughts and floods? Evidence from a discrete choice experiment in Odisha. *India. Water Resour. Econ.* **25**, 27–41 (2019).
106. Dar, M. H. *et al.* Drought tolerant rice for ensuring food security in eastern India. *Sustainability* **12**, 6 (2020).
107. Nayak, S. *et al.* Adoption trend of climate-resilient rice varieties in Bangladesh. *Sustainability* **14**, 5156 (2022).
108. Ahmed, B. *et al.* Indigenous people's responses to drought in northwest Bangladesh. *Environ. Dev.* **29**, 4 (2019).
109. Shayanmeh, S. *et al.* The impacts of climate change on water resources and crop production in an arid region. *Agriculture* **12**, 1056 (2022).

110. Easterling, D. R. *et al.* Climate extremes: observations, modeling, and impacts. *Science* (80-) **289**, 2068–2074 (2000).
111. Verdon-Kidd, D. C. & Kiem, A. S. Nature and causes of protracted droughts in southeast Australia: Comparison between the federation, WWII, and big dry droughts. *Geophys. Res. Lett.* **36**, 22 (2009).
112. Choruma, D. J., Akamagwuna, F. C. & Odume, N. O. Simulating the impacts of climate change on maize yields using EPIC: A case study in the Eastern Cape Province of South Africa. *Agriculture* **12**, 794 (2022).
113. Auffhammer, M., Hsiang, S. M., Schlenker, W. & Sobel, A. Using weather data and climate model output in economic analyses of climate change. *Rev. Environ. Econ. Policy* (2013).
114. Assunção, J. & Chein, F. Climate change and agricultural productivity in Brazil: Future perspectives. *Environ. Dev. Econ.* **21**, 581–602 (2016).
115. Gettliffe, E. *UN OCHA Anticipatory Action. Lessons from the 2020 Somalia Pilot*. <https://reliefweb.int/report/somalia/un-ocha-anticipatory-action-lessons-2020-somalia-pilot> (Center Disaster Protection London, 2021).
116. Kassie, B. T. *Climate Variability and Change in Ethiopia : Exploring Impacts and Adaptation Options for Cereal Production* (2014).
117. Hao, Z., Singh, V. P. & Xia, Y. Seasonal drought prediction: Advances, challenges, and future prospects. *Rev. Geophys.* **56**, 108–141 (2018).
118. Notton, G. & Voyant, C. *Chapter 3—Forecasting of Intermittent Solar Energy Resource*. (eds. Yahyaoui, I. B. T.-A. in R. E. and P. T.). 77–114. <https://doi.org/10.1016/B978-0-12-812959-3.00003-4> (Elsevier, 2018).
119. Jiang, D. & Wang, K. The role of satellite-based remote sensing in improving simulated streamflow: A review. *Water* **11**, 23 (2019).
120. Madadgar, S. *et al.* A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the south-western United States. *Water Resour. Res.* **52**, 34 (2016).
121. Slater, L. *et al.* Hybrid forecasting: Blending climate predictions with AI models. *Hybrid Forecast*. <https://doi.org/10.5194/hess-27-1865-2023> (2023).
122. Breiman, L. & Freedman, D. How many variables should be entered in a regression equation?. *J. Am. Stat. Assoc.* **78**, 131–136 (1983).
123. Breiman, L. & Ihaka, R. Nonlinear discriminant analysis via scaling and ACE. In *Technical Report No. 40* (1984).
124. Therneau, T., Atkinson, B. & Ripley, B. *Recursive Partitioning and Regression Trees* (2022).
125. Uddin, S., Khan, A., Hossain, M. E. & Moni, M. A. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inform. Decis. Mak.* **19**, 281 (2019).
126. Noble, W. S. What is a support vector machine?. *Nat. Biotechnol.* **24**, 1565–1567 (2006).
127. Kober, J., Bagnell, J. A. & Peters, J. Reinforcement learning in robotics: A survey. *Int. J. Rob. Res.* **32**, 1238–1274 (2013).
128. Freedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
129. Guelman, L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Syst. Appl.* **39**, 3659–3667 (2012).
130. Ridgeway, G. *gbm—Generalized Boosted Models*. R Package. 1–15 (2017).
131. Annette J. Dobson, A. G. B. *An Introduction to Generalized Linear Models*. <https://doi.org/10.1201/9781315182780> (Chapman and Hall/CRC, 2018).
132. Hothorn, T., Hornik, K. & Zeileis, A. Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Stat.* **15**, 651–674 (2006).
133. Seeger, D. E. & Hilt, D. W. *Ridge, a Computer Program for Calculating Ridge Regression Estimates* (1977).
134. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
135. Efron, B. *et al.* Least angle regression. *Ann. Stat.* **32**, 407–499 (2004).
136. Dey, D. K., Ghosh, S. K. & Mallick, B. K. *Generalized Linear Models: A Bayesian Perspective* (CRC Press, 2000).
137. Chen, T. Introduction to boosted trees. *Univ. Washington Comput. Sci.* **22**, 14–40 (2014).
138. Levshina, N. Conditional inference trees and random forests. In *A Practical Handbook of Corpus Linguistics*. 611–643 (Springer, 2020).
139. Sarda-Espinosa, A., Subbiah, S. & Bartz-Beielstein, T. Conditional inference trees for knowledge extraction from motor health condition data. *Eng. Appl. Artif. Intell.* **62**, 26–37 (2017).
140. Olusegun, A. M., Dikko, H. G. & Gulumbe, S. U. Identifying the limitation of stepwise selection for variable selection in regression analysis. *Am. J. Theor. Appl. Stat.* **4**, 414–419 (2015).
141. Friedman, J. H. & Roosen, C. B. An introduction to multivariate adaptive regression splines. *Stat. Methods Med. Res.* **4**, 197–217 (1995).
142. Milborrow, S., Hastie, T. & Tibshirani, R. *Earth: Multivariate Adaptive Regression Spline Models. R Package Version*. Vol. 3 (2014).
143. Ripley, B. D. *Pattern Recognition and Neural Networks* (Cambridge University Press, 2007).
144. Farikha, E. F., Hadi, A. F., Anggraeni, D. & Riski, A. Projection pursuit regression in statistical downscaling model using artificial neural network for rainfall prediction. *J. Phys. Conf. Ser.* **1872**, 12021 (2021).
145. Wright, M. N., Wager, S. & Probst, P. *Ranger: A Fast Implementation of Random Forests. R Package version 0.12*. Vol. 1 (2020).
146. Li, R. & Peng, L. Quantile regression for left-truncated semicompeting risks data. *Biometrics* **67**, 701–710 (2011).
147. Bello, R. W., Mohamed, A. S. A., Talib, A. Z., Sani, S. & Ab Wahab, M. N. Behavior recognition of group-ranching cattle from video sequences using deep learning. *Indian J. Anim. Res.* **56**, 505–512 (2022).
148. Ma, H., Li, T., Zhu, H. & Zhu, Z. Quantile regression for functional partially linear model in ultra-high dimensions. *Comput. Stat. Data Anal.* **129**, 135–147 (2019).
149. Ciner, C., Lucey, B. & Yarovaya, L. Determinants of cryptocurrency returns: A LASSO quantile regression approach. *Financ. Res. Lett.* **49**, 102990 (2022).

## Acknowledgements

The authors express their heartfelt gratitude to the Bangladesh Meteorological Department (BMD) for supplying the pertinent weather data.

## Author contributions

M.A.A.M.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing—original draft. M.R.S.: Resources; Supervision; Investigation, Validation; Visualization, Roles/Writing—original draft. M.A.R.S.: Resources; Supervision; Investigation, Validation; Visualization, Roles/Writing—review & editing. S.K.R.: Conceptualization; Methodology; Visualization; Roles/Writing—original draft; Data wrangling. S.A.I.N.: Roles/Writing—original draft. A.M.M.: Investigation, Supervision, Roles/Writing—review & editing. M.I.H.: Investigation, Supervision, Roles/Writing—review & editing. M.S.K.: Investigation, Supervision, Roles/Writing—review & editing. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-51111-2>.

**Correspondence** and requests for materials should be addressed to M.A.R.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024