



# Impact of climate and weather extremes on soybean and wheat yield using machine learning approach

Mamta Kumari<sup>1</sup> · Abhishek Chakraborty<sup>1</sup> · Vishnubhotla Chakravarathi<sup>2</sup> · Varun Pandey<sup>1</sup> · Parth Sarathi Roy<sup>3</sup>

Accepted: 11 June 2024 / Published online: 13 July 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

The escalating climate instability and extreme weather events significantly jeopardize food security. The study assessed the impact of long-term climatic variables and extreme weather events on soybean and wheat yields in rainfed central India. To address inherent spatial variability, the study area was divided into homogeneous zones based on rainfall and soil parameters. Crop yields were correlated with a comprehensive set of driving variables at seasonal and monthly scales within each zone. Machine learning algorithms, including Random Forest Regression (RFR) and Neural Networks (NN), were employed to analyze crop yield anomalies caused by climate and weather extremes. The Sobol' index was utilized for global sensitivity analysis to identify key parameters. Results showed significant negative correlations between thermo-meteorological parameters and yields of both monsoon soybean and winter wheat across multiple districts. Soybean yield exhibited a notable positive correlation with hydro-meteorological parameters, while wheat yield displayed a significant positive correlation with cold temperature extremes. RFR and NN demonstrated similar performance, with Root Mean Square Error (RMSE) values ranging from 0.27 to 0.39 t/ha for soybean and 0.4 to 0.6 t/ha for wheat. The Sobol' index highlighted the high sensitivity of soybean yield to rainfall and rainy days during July and August, corresponding to the crop development and flowering stages. In contrast, wheat yield was primarily influenced by temperature extremes, particularly cold nights and hot days during the reproductive-maturity stage. These crop- and growth-stage-specific analyses of meteorological parameters are essential for devising effective strategies to adapt and mitigate climate emergencies.

**Keywords** Crop yield · Central India · Random forest · Neural network · Sobol' index

## 1 Introduction

The Intergovernmental Panel on Climate Change's (IPCC) Sixth Assessment Report (2021) emphasizes persistent challenges in climatic variables, including rising temperatures, erratic rainfall patterns, and an increasing frequency of extreme weather events across South Asia. Significantly,

this region is inhabited by over 70% of small farmholders, who are notably susceptible to the effects of extreme weather events (Harvey 2018). Furthermore, the impacts of climate change on crop yield demonstrate spatiotemporal variations, intensifying concerns regarding food security (Praveen and Palanivelu 2017). Overall, the IPCC report highlights the paramount importance of mitigation and adaptation strategies in addressing the influence of climate change on crop yields and sustainable agriculture. This necessitates comprehensive understanding of the interaction between climatic variables/weather extremes and crop yields (Webber et al. 2018).

While long-term climatic variables determine the regional crop growth potential, their impact on crop yields varies significantly by the influence of weather extremes. Weather extremes can result in yield reduction below region-specific potential, persisting even when crops are exposed to optimal weather conditions (Tigchelaar et al. 2018). In general, process-based and statistical approaches

✉ Mamta Kumari  
mamta9507@gmail.com

<sup>1</sup> Agroecosystem and Modeling Division, Agriculture Sciences and Application Group, Remote Sensing Applications Area, National Remote Sensing Centre, Balanagar, Hyderabad, Telangana 500037, India

<sup>2</sup> Centre for Earth, Ocean and Atmospheric Sciences, University of Hyderabad, Gachibowli, Telangana 500032, India

<sup>3</sup> Sustainable Landscapes and Restoration, World Resources Institute of India, New Delhi 110016, India

are being employed to assess and quantify climate change-induced crop yield losses. Several studies have described the efficacy of both approaches in discerning the criticality of climate and weather extremes influencing crop yields and decision-making (Lobell and Asseng 2017). Process-based crop models (De Wit 1965) simulate crop growth and development but are incapable of generating historical crop yield anomalies. Additionally, these models necessitate extensive and often hard-to-obtain input parameters (Van Oort et al. 2011; Lu et al. 2017), which are frequently omitted from the modeling process, thus compromising accuracy (Robert et al. 2017). In contrast, statistical modeling encompasses the effects of climate and weather extremes on crop yield by utilizing historical data (Moore et al. 2017; Kukul and Irmak 2018) through two approaches: linear/non-linear parametric and non-linear non-parametric (or machine learning) statistical models. These models can predict crop yield under varying weather conditions, including rainfall, temperature, drought indices, etc. Recently, machine learning (ML) techniques such as random forest and neural networks have been employed for impact assessment and yield estimation (Crane-Droesch 2018; Vogel et al. 2019; Konduri et al. 2020; Schierhorn et al. 2021). Lu et al. (2017) suggested that statistical models require fewer resources compared to process-based models. Furthermore, these can account for spatial heterogeneities in yield, climate, and soil conditions, making ML models the preferred choice for performance.

Wheat, a dominant cereal crop, and soybean, a legume-oilseed crop, play crucial roles in ensuring food and nutritional security, but are significantly affected by recent climatic fluctuations. Wheat constitutes 36% of global food grain production, with Central India being a major contributor as per CIMMYT, Mexico (Braun et al. 1997). It is highly sensitive to temperature increase, with yield declining significantly above 34 °C due to accelerated senescence (Lobell et al. 2012). Temperature requirements vary across growth stages, such as 22 °C for the vegetative stage, 21 °C for the reproductive stage, and 35 °C as the maximum limit for the grain-filling stage (Porter and Gawith 1999), making the effects of extreme weather events time-specific (Powell and Reinhard 2016). Soybean contributes to 25% of global edible oil production and is crucial for livestock feed. Yield declines of soybean occurs outside the optimal temperature range of 26 °C to 30 °C, with adverse effects on flowering and pod setting. Insufficient precipitation negatively impacts soybean yield, with an ideal rainfall range of 500–1000 mm (MacCarthy et al. 2022). Though soybean can grow with as little as 180 mm of rain in a season, it results in a 40 to 60% yield reduction compared to optimal conditions. Considering the sensitivity of crop growth stages to temperature and rainfall, the inclusion of monthly mean climatic data,

in conjunction with seasonal data, is a valuable addition to the analysis.

The influence of climatic variables and weather extremes on crop productivity is intricate, influenced by various factors such as geographical location, crop type, and prevailing agricultural practices. For instance, Kang et al. (2009) noted that regions characterized by high soil water holding capacity exhibit resilience to climate variability impacts while sustaining crop yields. Recent research has endeavored to assess the influence of weather and climate on crop yield; however, much of it has relied on regression models to establish relationships between climatic variables such as temperature and rainfall and crop yield (Novikova et al. 2020; Paymard et al. 2018; Zampieri et al. 2017; Matiu et al. 2017; Yu et al. 2013). However, there is a growing recognition of the need to explore further into this relationship by considering additional factors such as weather extremes during crop-specific developmental phases in different agro-ecological conditions. While a few studies have employed machine learning (ML) techniques to capture the complex nonlinear associations between yield and climatic variables, the incorporation of weather extremes remains limited. For instance, Peichl et al. (2018) and Hofman et al. (2020) explored crop-yield dependencies on crop-specific development phases using ML models but did not include extreme weather events or conduct agro-ecological area-specific evaluations. Beillouin et al. (2020) examined the impact of diverse extreme weather and climatic conditions on historical yield anomalies in European crop production using ML techniques but did not consider crop development phases. Similarly, Schierhorn et al. (2021) investigated the contribution of climate and weather extremes to wheat yield across various growth stages but did not account for agro-ecological conditions.

In the Indian context, studies analyzing the impact of climate and weather extremes on crop yield using ML techniques and considering different development phases are scarce. While some research, such as that by Mohanty et al. (2017) and Birthal et al. (2014), has explored the impact of climatic scenarios on soybean and other crop yields, these studies often focus on temperature and rainfall, neglecting the influence of extreme weather events. Mohapatra et al. (2022) found that increasing temperatures significantly affected grain yield in eastern India, while Gupta et al. (2022) and Madhukar et al. (2021) highlighted adverse effects of temperature and rainfall anomalies on crop yields, with limited attention to extreme weather events. Few investigations have focused on explaining the impact of heat waves on wheat yield, often confined to single-year analyses (Chakraborty et al. 2019) or employing linear regression approaches (Rao et al. 2015). None of these studies have comprehensively addressed the full range of climatic

factors and their associated extremes. Furthermore, the Sobol’ index, a widely utilized global sensitivity analysis method, is employed in the present study to identify sensitive parameters by simultaneously varying them across the entire parameter space (Sobol 1993). Therefore, understanding the sensitivity of soybean and wheat yields to climatic variability and identifying the critical climate variables affecting these crop yields in India is imperative for effective agricultural management strategies. This study utilized ML models to discern vital climatic factors (monthly and seasonal) linked to yield anomalies in soybean and wheat in homogenous zones. Our findings contribute to existing knowledge by identifying the climatic variables specific to each growth stage that had the most significant impact on soybean and wheat yield losses in India, based on long-term historical data.

## 2 Study area

Two predominantly rainfed states in India, Maharashtra and Madhya Pradesh, both of which have experienced several drought events in recent past and are known for cultivating soybean and wheat, were chosen as the study areas (Fig. 2).

Maharashtra, situated in western India, extends from latitudes 15°35’ N to 22°02’ N and longitudes 72°36’ E to 80°54’ E, covering an area of 3.08 lakh square kilometers, and divided into 36 districts. In contrast, Madhya Pradesh, the second-largest state in India, encompasses 50 districts which spans from latitudes 17°48’ N to 26°52’ N and longitudes 74°2’ E to 84°24’ E. Both states experience a monsoon season from June to September, with higher rainfall in the south and southeast compared to the northwest. Approximately 55% of Maharashtra and 50% of Madhya Pradesh’s total land area is utilized for cultivation. Soybean and Wheat are the major *kharif* (monsoon) season and *rabi* (winter) season crop in Maharashtra and Madhya Pradesh, respectively. The crop calendar of Soybean and Wheat crops, in respective study area, is given in Table 1.

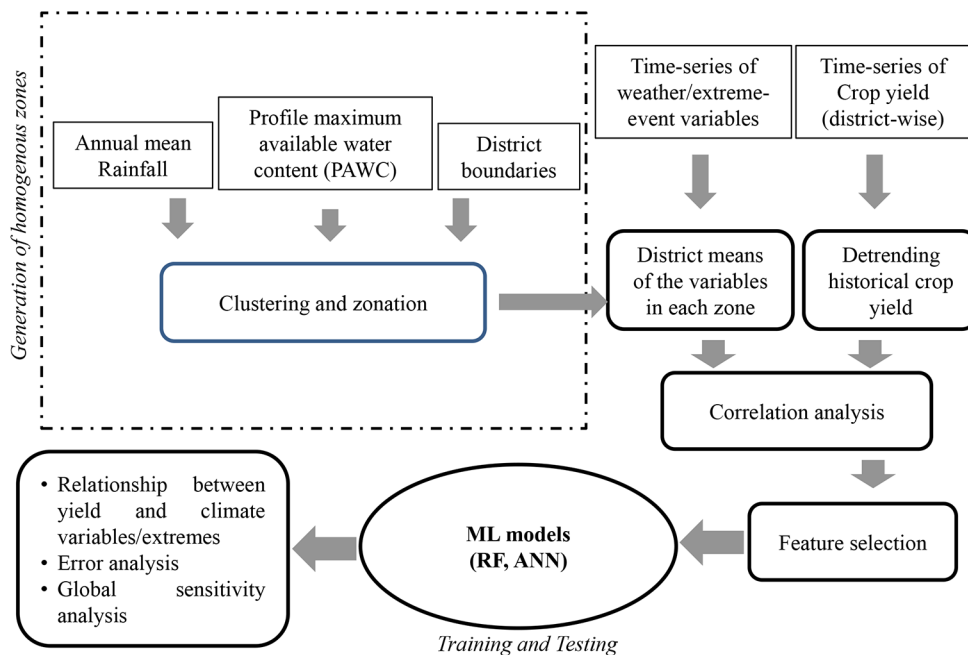
## 3 Data used and methodology

The schematic diagram of methodology is presented in Fig. 1.

**Table 1** Crop calendar of Soybean and Wheat grown in *kharif* and *rabi* season, respectively

| State          | Crops   | Initial | Crop development | Reproductive | Maturity | Harvest |
|----------------|---------|---------|------------------|--------------|----------|---------|
| Maharashtra    | Soybean | Jun-Jul | Jul-Aug          | Aug-Sep      | Sep-Oct  | Oct     |
| Madhya Pradesh | Wheat   | Nov-Dec | Dec-Jan          | Jan-Feb      | Feb-Mar  | Mar     |

**Fig. 1** Schematic diagram of methodology



### 3.1 Clustering of the study area into homogeneous zones

Both Maharashtra and Madhya Pradesh exhibit significant spatial variability in terms of climate and soil types. Consequently, agricultural practices in these states vary considerably across different regions, which is evident in the spatial distribution of crop types and crop calendars. In this study, the impact of climatic and weather extremes on wheat and soybean yields was studied by delineating the study states

into homogeneous zones (Fig. 2). This division was based on rainfall patterns and Profile Available Water Capacity (PAWC) of soil using a rule-based approach. PAWC refers to the capacity of the soil to store water in the soil profile as available soil moisture. It comprises water retained between field capacity and permanent wilting point of soil. Factors influencing PAWC include soil texture, organic matter content and depth. Both rainfall and PAWC are critical factors in crop production, as the former affects water supply while the latter determines water availability for crop growth.

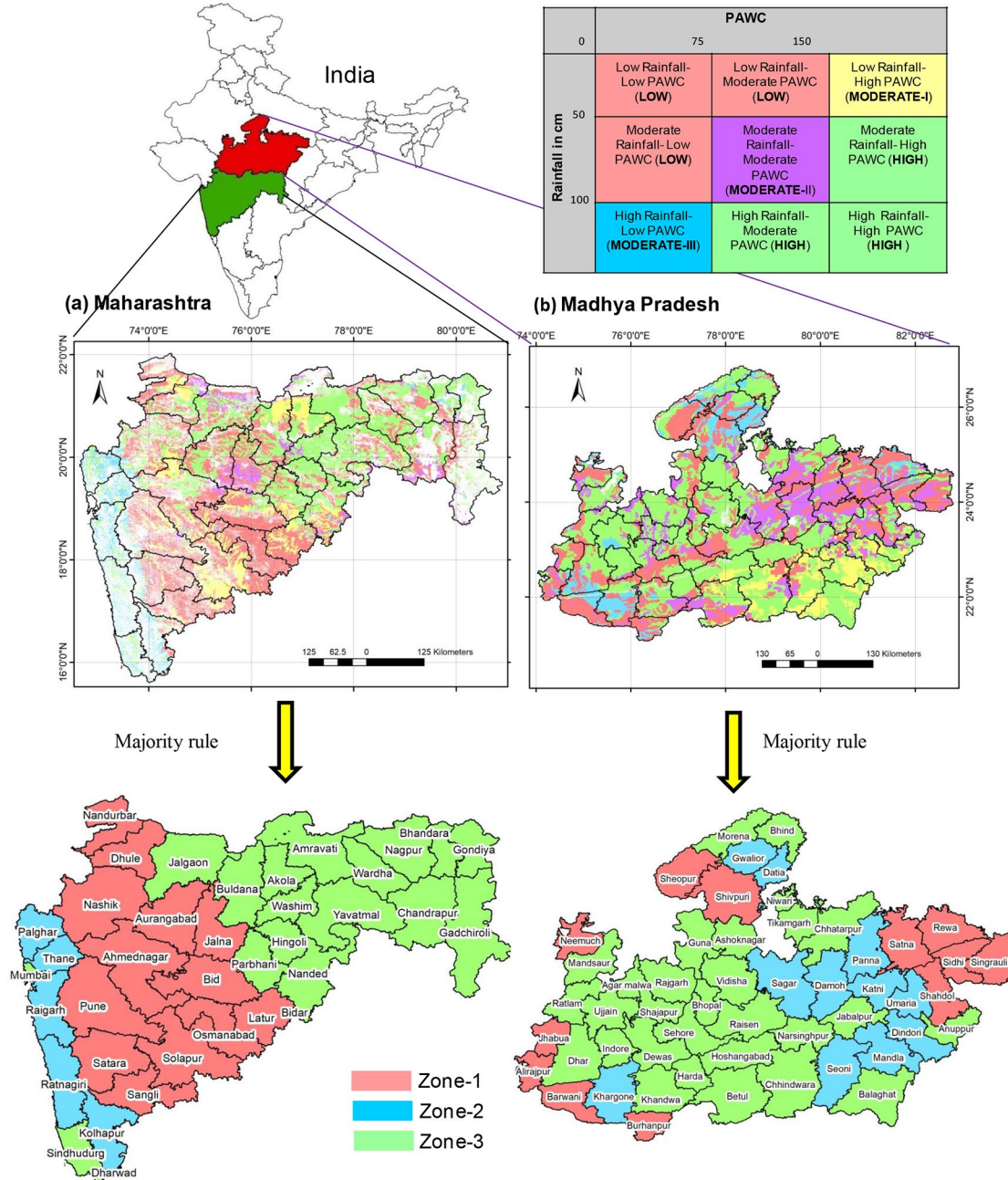


Fig. 2 Homogenous zones delineation in study area using rainfall and PAWC



**Table 2** Characteristics of various delineated zones in study area

| State                  | Zones   | % Area | Climate              | Soil                            | LGP (days) |
|------------------------|---|--------|----------------------|---------------------------------|------------|
| Maharashtra (Soybean)  | Zone-1: Rainfall ( $\leq 100$ cm)/ PAWC ( $\leq 150$ cm): both are limiting     | 48.87  | Hot semi-arid        | Shallow/medium/ deep black soil | 90–150     |
|                        | Zone-2: Either rainfall $< 100$ cm or PAWC $< 150$ cm: One of these is limiting | 20.5   | Hot humid/ Per-humid | Red-lateritic and alluvium soil | $> 210$    |
|                        | Zone-3: Rainfall ( $> 100$ cm)/ PAWC ( $> 150$ cm): both are sufficient         | 38.05  | Hot sub-humid        | Red and black soil              | 150–180    |
| Madhya Pradesh (Wheat) | Zone-1: Rainfall ( $\leq 100$ cm)/ PAWC ( $\leq 150$ cm): both are limiting     | 22.6   | Hot semi-arid        | Alluvium-derived soils          | 90–150     |
|                        | Zone-2: Either rainfall $< 100$ cm or PAWC $< 150$ cm: One of these is limiting | 23.3   | Hot semi-arid        | medium/ deep black soil         | 90–150     |
|                        | Zone-3: Rainfall ( $> 100$ cm)/ PAWC ( $> 150$ cm): both are sufficient         | 55.1   | Hot sub-humid        | Red and black soil              | 150–180    |

Annual mean rainfall data of CHIRPS ( $5 \text{ km} \times 5 \text{ km}$ ) and PAWC derived from soil map of National Bureau of Soil Survey & Land Use Planning (NBSS&LUP) at 1:50,000 scales was employed to divide Maharashtra and Madhya Pradesh into homogeneous zones.

Initially, rainfall and PAWC were categorized into three classes for each parameter across the study area, as mentioned in Fig. 2. Subsequently, by combining the classes from each parameter, a total of nine classes were generated, forming a  $3 \times 3$  matrix. Majority rule was applied to assign a single class to each district (Fig. 2). As a result, each state was delineated into three homogeneous zones, grouping together districts with almost similar characteristics. Salient characteristics of the homogenous zones of Maharashtra and Madhya Pradesh were taken from the respective agro-ecological zones and are presented in Table 2 (Subramaniam 1983).

### 3.2 Predictor variables employed in the analysis

In this study, a range of predictor variables were obtained, systematically organized, and used to compute climatic parameters and weather extremes spanning the period from 1999 to 2019, as outlined in Table 3. During the *kharif* season, predictor variables for four consecutive months (June, July, August, and September) were employed to conduct an impact analysis of soybean cultivation in Maharashtra. Similarly, during the *rabi* season, predictor variables spanning five months (November, December, January, February, and March) were utilized to evaluate their impact on

**Table 3** Predictor variables used (1999–2019) for studying the impact of mean and extreme weather on crop yield

| Predictor variables   | Definition   | Source  |
|---|--|---|
| Tmax, Tmin (monthly)  | Monthly Mean of maximum temperature and minimum temperature during the growing season  | CPC temperature ( $0.5^\circ \times 0.5^\circ$ ): <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html</a> |
| Rainfall and Rainy days (monthly/seasonal)                          | Sum of monthly rainfall during the growing season, no. of days in season when Rainfall $> 2.5$ mm  | CHIRPS ( $0.05^\circ \times 0.05^\circ$ ): <a href="https://data.chc.ucsb.edu/products/CHIRPS-2.0/">https://data.chc.ucsb.edu/products/CHIRPS-2.0/</a>  |
| Growing Degree Days (GDD): monthly and seasonal                     | Sum of difference of daily mean temperature and reference temperature ( $10^\circ\text{C}$ for Soybean/ $5^\circ\text{C}$ for Wheat) in growing season                                   | CPC temperature ( $0.5^\circ \times 0.5^\circ$ ): <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html</a> |
| Standardized Precipitation Evapotranspiration index (SPEI): monthly | Time-series of the climatic water balance (P-PET) fitted to log-logistic and then Gaussian distribution to get SPEI. Here, SPEI at 3-months time-interval during growing season are used | CRU ( $0.5^\circ \times 0.5^\circ$ ): <a href="https://spei.csic.es/database.html">https://spei.csic.es/database.html</a>   |
| Maximum Temperature (Tmax) extremes (Hot day, Cold day): monthly    | No. of days during the growing season with daily Tmax $> 90^{\text{th}}$ percentile (Hot day); with daily Tmax $< 10^{\text{th}}$ percentile (Cold day)                                  | CPC temperature ( $0.5^\circ \times 0.5^\circ$ ): <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html</a> |
| Tmin extremes (Hot night, cold night)                               | No. of days during the growing season with daily Tmin $> 90^{\text{th}}$ percentile (Hot night); with daily Tmin $< 10^{\text{th}}$ percentile (Cold night)                              | CPC temperature ( $0.5^\circ \times 0.5^\circ$ ): <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html</a> |

wheat cultivation in Madhya Pradesh. In total, 43 variables were computed using daily data for soybean crop analysis, while 53 variables were computed for wheat crop analysis. Soybean and wheat crop yield data were obtained from the Ministry of Agriculture website (<https://data.desagri.gov.in/website/crops-apy-report-web>) for the Maharashtra and Madhya Pradesh regions, respectively. To remove inherent trends in the data, a linear regression detrending approach was applied for crop yield data preprocessing.

### 3.3 Correlation between yield and climatic parameters/ extremes

A Pearson's correlation coefficient was calculated to examine the relationship between crop yields and the various weather and climatic variables. This provides insight into how climatic and weather extremes affect crop yields. The correlation study was carried out on seasonal and monthly basis to understand the specific impact of weather extremes on crop yield during different growth stages. It is important to note that correlation does not imply causation, so the study was followed by more detailed analysis of the data. The formula to compute the correlation coefficient ( $r$ ) is given below:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (1)$$

Where,  $x_i$  is the crop yield in time-series data in a point  $i$  for a particular district,  $y_i$  is the weather/climatic variables in time-series data in a point  $i$  for a particular district,  $\bar{x}$  and  $\bar{y}$  is district mean yield and weather/climatic variables in time-series, respectively, and  $N$  is total number of data points.

### 3.4 Development of models

For the present study, two ML models viz., Random Forest regression (RFR) and Neural network (NN) are developed for each zone and each crop. Approximately, 70% of data were used for model training and remaining 30% were used for model testing. For each model, a grid-search and 10-fold cross validation technique was used to select the optimum hyperparameters to tune the machine learning models. Then, for each grid location, performance of each model was evaluated using statistical measures of error between the predicted and observed values i.e., Root Mean Squared Error (RMSE). A well-trained statistical model generates small RMSE values. The optimum values of hyper-parameters producing maximum accuracy for each ML model are given in Table 4. District-level seasonal and month-wise

climatic, weather extreme variables are predictor variables in this analysis, while the crop yields (soybean and wheat) as response variable in each zone.

#### 3.4.1 Random forest regression (RFR) model

Random Forest Regression is an ensemble machine learning algorithm (Breiman 2001) used for regression (or classification) tasks, able to handle a large volume of input data and simultaneously can reduce overfit. RFR is composed of multiple decision trees (DT), and the final prediction is an aggregation of the predictions made by these individual trees. The accuracy of the random forest algorithm relies mainly on the strength of the individual tree classifiers and the dependency between the classifiers (Amit and Geman 1997). The inclusion of several trees increases the probability of deriving an effective prediction model (Breiman 2001; Strobl et al. 2008). The key components and steps involved in RFR with input features ( $x$ ) and corresponding target values ( $y$ ) are given below:

- RFR consists of an ensemble of DT. Each tree is trained independently on a subset of the data using a process called bootstrapped sampling (randomly selecting samples with replacement) as per Liaw and Wiener (2002).
- Consider a dataset of size  $P \times Q$ , where  $P$  is the number of data points and  $Q$  is the number of features.

Let  $P1$  data points be present at some node  $d$ . RFR selects  $K$  random features out of  $Q$  ( $K < Q$ ); and for each feature net reduction in variance is computed. The feature for which net reduction in variance is maximum is selected for decision making.

The criteria for nodal feature selection can be expressed as follows:

$$f_{selected\ feature} = \operatorname{argmax}_i \left( \Delta V_d^{f_i} \right) \quad (2)$$

$$\Delta V_d^{f_i} = V_d - (\alpha_1 V_{S1}^{f_i} + \alpha_2 V_{S2}^{f_i}) \quad (3)$$

$$\alpha_1 = \frac{\text{Number of data points in } S1}{P1} \quad (4)$$

$$\alpha_2 = P1 - \alpha_1 \quad (5)$$

Where:

$V_d$  is the variance at node  $d$ ,

$\Delta V_d^{f_i}$  represents the reduction in the variance at node  $d$  when  $f_i$  is selected for decision making,

$V_{S1}^{f_i}$  is the variance present in  $S1$  subset when feature  $f_i$  is selected for decision making ( $i \in (1, K)$ ), and

**Table 4** Hyperparameters range and the optimal values for each ML model for yield modeling

|                              | Model description                | Parameters           | Zones                           |                      |                      |                      |     |
|------------------------------|----------------------------------|----------------------|---------------------------------|----------------------|----------------------|----------------------|-----|
|                              |                                  |                      | Parameters range with step size | Zone-01              | Zone-02              | Zone-03              |     |
|                              |                                  |                      |                                 | Optimized parameters | Optimized parameters | Optimized parameters |     |
| Soybean                      | RFR                              | Number of trees      | 10 to 500 @ 1                   | 274                  | 219                  | 351                  |     |
|                              | Loss function: RMSE              | Minimum sample split | 2 to 8 @ 1                      | 2                    | 2                    | 3                    |     |
|                              | Maximum no. of feature: Sqrt     | Minimum sample leaf  | 1 to 6 @ 1                      | 1                    | 1                    | 2                    |     |
|                              | NN Learning function: Leaky ReLU | Learning rate        | 0.001 to 1 @ 0.01               | 0.01                 | 0.01                 | 0.02                 |     |
|                              | Learning algorithm: SGD          | momentum             | 0.1 to 1 @ 0.1                  | 0.8                  | 0.9                  | 0.8                  |     |
|                              | Loss function: RMSE              | Batch size           | 10 to 100 @ 10                  | 30                   | 30                   | 30                   |     |
|                              |                                  | training epochs      | 50 to 300 @10                   | 120                  | 190                  | 150                  |     |
|                              | Wheat                            | RFR                  | Number of trees                 | 10 to 500 @ 1        | 318                  | 252                  | 290 |
|                              |                                  | Loss function: RMSE  | Minimum sample split            | 2 to 8 @ 1           | 2                    | 2                    | 3   |
| Maximum no. of feature: Sqrt |                                  | Minimum sample       | 1 to 6 @ 1                      | 1                    | 2                    | 2                    |     |
| NN Learning function: ReLU   |                                  | Learning rate        | 0.001 to 1 @ 0.01               | 0.02                 | 0.01                 | 0.02                 |     |
| Learning algorithm: Adam     |                                  | momentum             | 0.1 to 1 @ 0.1                  | 0.5                  | 0.6                  | 0.8                  |     |
| Loss function: RMSE          |                                  | Batch size           | 16 to 256 @ 10                  | 20                   | 20                   | 20                   |     |
|                              |                                  | training epochs      | 10 to 300 @10                   | 200                  | 210                  | 200                  |     |

$V_{S_2}^{fi}$  is the variance present in  $S_2$  subset when feature  $fi$  is selected for decision making ( $i \in (1, K)$ ).

- Each of the trees is grown till maximum depth to reduce bias. Finally, the RFR aggregates the predictions from all the individual decision trees by averaging, which reduces the overall variance of model (Breiman 2001).

Mathematically, the prediction made by a RFR for regression can be represented as follows:

Given a new data point with feature values  $x_{new}$ , and a RFR with N decision trees, the prediction ( $y_{pred}$ ) is:

$$y_{pred} = \frac{\sum_{i=1}^N (tree_i(x_{new}))}{N} \tag{6}$$

Where,  $tree_i(x_{new})$  represents the prediction of the  $i^{th}$  decision tree for the new data point  $x_{new}$ .

### 3.4.2 Neural network (NN) model

A neural network model with 2 hidden layers (input-8-4-1) was constructed to predict the zone-wise yield of soybean and wheat using Python programming language and PyTorch library (Paszke et al. 2019). Several hyperparameters (number of hidden layers, number of neurons per layer, learning rate, and batch size) which affect the performance of an ANN are tuned using grid-search optimization and 10-fold cross validation technique (Table 4). Here, data was used in a 70:30 ratios to train and test the model. The leaky Rectified Linear Unit (ReLU) activation function was used in each layer with decay coefficient of 0.01. The model was optimized using momentum-based Stochastic Gradient Descent (SGD) optimizer algorithm. Momentum-based SGD is a variation of the traditional SGD optimization algorithm that incorporates momentum to accelerate convergence. The momentum term allows the optimizer to build up velocity in directions where gradients consistently point, which accelerates convergence and escape local minima. The Mean Square Error (MSE)

was used as loss function to evaluate the model. The steps involved in NN training are as follows:

1. Initialize model parameters (weights and biases) with Xavier initialization (Glorot and Bengio 2010)
2. Define Leaky ReLU as activation function as follows

$$\text{Leaky ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha^* x, & \text{if } x \geq 0 \end{cases} \quad (7)$$

3. Define a list of learning rate, batch size and momentum parameters to be tested for model performance.
4. For each combination in step 3, train the NN model and update model parameters via SGD with momentum (Eqs. 8 and 9) and evaluate model performance on validation dataset.

$$W_{t+1} = W_t + V_{t+1} \quad (8)$$

$$V_{t+1} = \mu \times V_t - \alpha \times \nabla L_t \quad (9)$$

Where  $W_t$  is weight at  $t^{\text{th}}$  epoch,

$V_t$  is velocity at  $t^{\text{th}}$  epoch,

$\mu$  is decaying factor accounts the number of iterations of previous gradients into the current update;  $\mu \in (0, 1)$ .

$\alpha$  is learning rate,

$\nabla L_t$  is gradient of loss function at  $t^{\text{th}}$  epoch.

5. Compute loss for each of the model-parameter combination in step 3 and select the model parameter with corresponding to minimum MSE.

### 3.5 Evaluation of modeling performance

Three statistical criteria, including coefficient of determination ( $R^2$ ), root mean square error (RMSE), and normalized root mean square error (nRMSE), are used to evaluate the performance of the ML models. Equations to compute RMSE and nRMSE are given below:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y})^2}{N}} \quad (10)$$

$$\text{nRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y})^2}{N}}}{\bar{y}} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

Where,  $y_i$  is the actual crop yield in time-series data in a point  $i$  for a particular district,  $\hat{y}$  is estimated yield using model in same district,  $N$  is total number of data points,  $\bar{y}$  is district mean yield in time-series.

### 3.6 Impact analysis of predictive variables in crop yield estimation

Impact of various predictor variables on crop yield estimation was analyzed through sensitivity analysis under different presumed scenarios. We assumed the accuracy to be highest with least RMSE, with all predictor variables. Therefore, we systematically removed individual predictor variables to evaluate their influence on crop yield estimation using different machine learning approaches. This analysis aims to quantify the significance of predictor variables in predicting crop yield based on historical data. Combinations of predictor variables under different scenarios were analyzed for soybean and wheat crops during their respective growing seasons. The following is a summary of the scenarios:

- Scenario 1 (S1): All predictor variables were retained in each machine learning model for both soybean and wheat crops.
- Scenario 2 (S2): Monthly temperature extremes are excluded from the input data.
- Scenario 3 (S3): Monthly minimum temperatures, maximum temperatures, and temperature extremes are excluded from the input data.
- Scenario 4 (S4): Both monthly and seasonal Growing Degree Days (GDD), as well as minimum temperatures, maximum temperatures, and temperature extremes are excluded from the input data.
- Scenario 5 (S5): Monthly and seasonal rainfall, rainy days, GDD, minimum temperatures, maximum temperatures, and temperature extremes are excluded from the input data.
- Scenario 6 (S6): Monthly SPEI, GDD, minimum temperatures, maximum temperatures, and temperature extremes are excluded from the input data; retaining monthly and seasonal rainfall and rainy days.

### 3.7 Global sensitivity analysis

In the current study, a global sensitivity index, often referred to as the first-order Sobol' index is used (Sobol 2001). It measures the first-order effect of an input variable on the output of a model and quantifies the portion of the total output variance that can be attributed to the variation in a single input variable while keeping all other input variables fixed.



Mathematically, the first-order Sobol’ index for an input variable  $x_i$  is defined as follows:

Let  $y$  be the output of the model, and let  $x_i$  be the  $i^{\text{th}}$  input variable. Then, the first-order Sobol’ index  $S_i$  is calculated as:

$$S_i = \frac{V[E(y|x_i)]}{V(y)} \tag{13}$$

Where,  $S_i$  is the first-order Sobol’ index for input variable  $x_i$ ,  $E(y|x_i)$  is the expected value (mean) of the output,  $y$  conditioned on the value of input variable  $x_i$ . In other words, it’s the average value of the output when  $x_i$  varies while all other inputs are fixed.  $V$  represents the variance of a random variable.  $V(y)$  is the variance of the total output  $y$ , considering all input variables.

The first-order Sobol’ index provides insight into how much the variation in an individual input variable contributes to the overall variability of the model output. A high

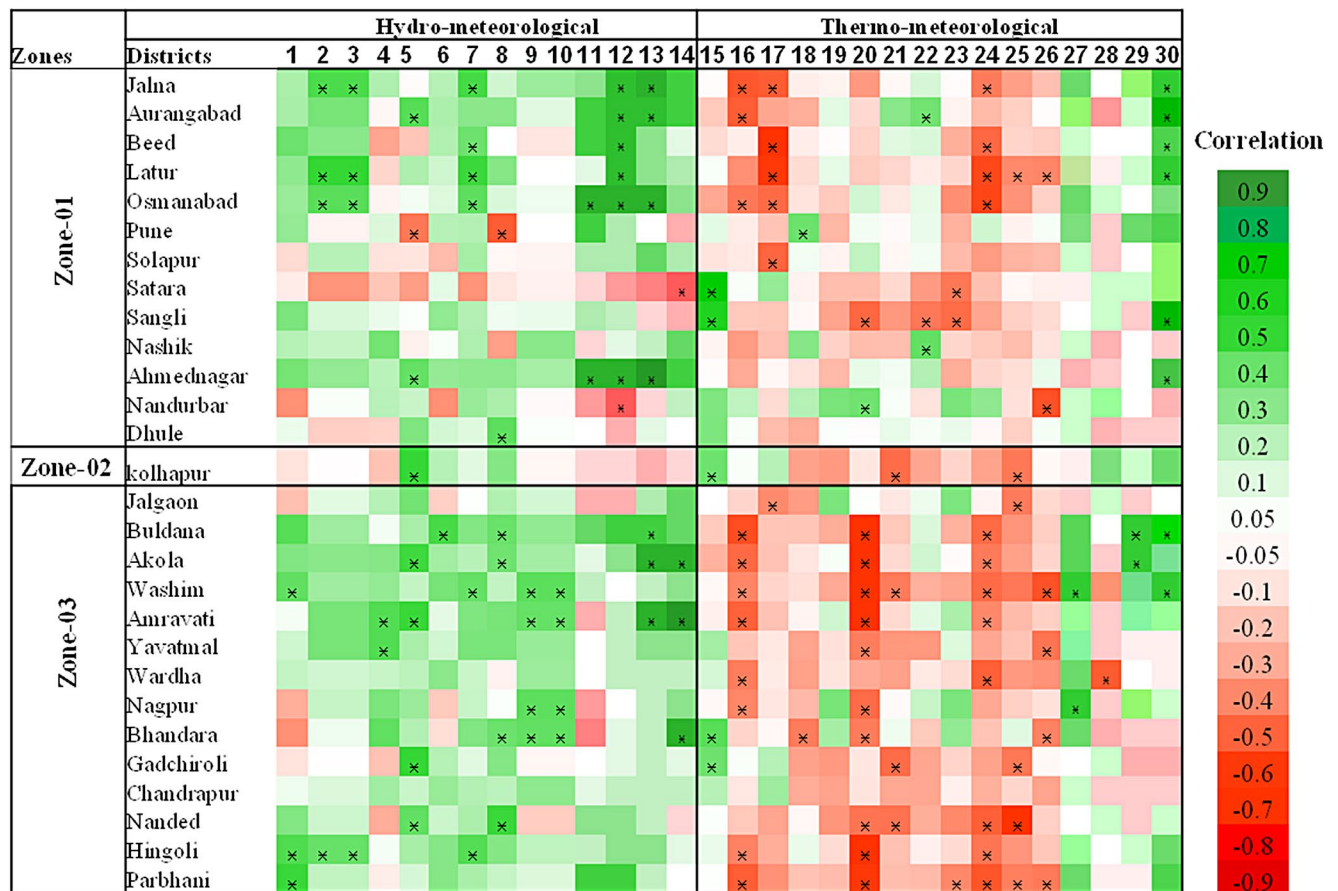
first-order Sobol’ index suggests that the variable has a significant influence on the output, while a low index suggests that the variable’s effect is relatively small.

## 4 Results

### 4.1 Correlation between crop yield and different predictor variables

Different hydro-meteorological (rainfall, rainy days, SPEI) and thermo-meteorological (Tmax, Tmin, GDD) parameters over monthly/ seasonal scale along with the temperature extremes (hot days, hot nights, cold days, cold nights) are linearly correlated with the district level crop yield over the time period of 1999–2019. Zone-wise correlation matrix for soybean and wheat yield is presented in Tables 5 and 6 respectively. As detailed in Table 2, in zone 1, both rainfall and Plant Available Water Capacity (PAWC) are limiting

**Table 5** Correlation coefficient between Soybean yield and climate/weather extremes over different zones of Maharashtra



\* Statistically significant at  $p < 0.1$

**Hydro-meteorological variables:** (1: RF\_Jun, 2:RF\_Jul, 3: RF\_Aug, 4: RF\_Sep, 5: RF Seasonal, 6: RD\_Jun, 7: RD\_Jul, 8: RD\_Aug, 9: RD\_Sep, 10: RD Seasonal, 11: SPEI\_Jun, 12:SPEI\_Jul, 13: SPEI\_Aug, 14: SPEI\_Sep)

**Thermo-meteorological variables:** (15: Tmax\_Jun, 16: Tmax\_Jul, 17: Tmax\_Aug, 18: Tmax\_Sep, 19: Tmin\_Jun, 20: Tmin\_Jul, 21: Tmin\_Aug, 22: Tmin\_Sep, 23: GDD\_Jun, 24:GDD\_Jul, 25:GDD\_Aug, 26: GDD\_Sep, 27: Cold days, 28: Cold nights, 29: Hot days, 30: Hot nights)



Similarly, the GDD of July, August, and September exhibited a significant negative correlation with soybean yield. However, temperature extremes did not show any significant correlation. It is apt to mention that soybean is cultivated in the warm and humid climate of the *kharif* season in central India. Chakraborty et al. (2017) reported a significant increase in temperature, particularly  $T_{min}$ , in these areas due to climate change. This rise in temperature has a cascading effect on the *kharif* season GDD, making it significantly higher, and consequently, negatively impacting soybean yield. The present study also revealed a significant negative correlation between temperature (particularly  $T_{max}$ ,  $T_{min}$ , and GDD) during the crop establishment phase and soybean yield.

As the majority of soybean crops are rainfed, hydro-meteorological parameters generally exhibit a positive correlation with soybean yield. Specifically, rainfall in July and August, rainy days in July, and the SPEI in July and August were significantly positively correlated with soybean yield across most districts in zone 1. In zone 3, seasonal rainfall, rainfall in September, seasonal rainy days, rainy days in August and September, and SPEI in August and September showed significant positive correlations with soybean yield.

In zone 1, low rainfall during the crop establishment and development phases may negatively affect yield, especially if the soil's water-holding capacity is insufficient to meet the crop's requirements. However, in zone 3, significant yield reductions were observed only during critical growth stages, such as flowering and pod filling, when water demand is high and rainfall is scarce. Therefore, rainfall and its distribution have a positive effect on soybean yield. Additionally, SPEI, which reflects the overall moisture surplus or deficit condition of the ecosystem, also exhibits a positive correlation with soybean yield.

In India, wheat (a long-day plant) is cultivated during the winter season to cope with rising temperatures during its crucial ripening and reproductive growth phases. Table 5 represents the correlations between climatic variables and their extremes with district-level wheat yield in three zones. As nearly half of the crop's water requirements are typically met by rainfall, a significant correlation was not observed in most districts in Madhya Pradesh, indicating the complexity of the relationship between rainfall patterns and wheat yield.

The SPEI of November, January, March, Rainy days over the season and November month showed significant positive correlation with wheat yield over Zone (1) Similarly, SPEI of February and March were found to have significant positive correlation with wheat yield over zone (2) Monthly SPEI of December to March and Rainfall of December showed similar result over zone (3) In nutshell, rainfall and moisture surplus condition as depicted by SPEI has significant positive effect on

wheat yield as it is grown under limited irrigation condition in Madhya Pradesh.

Effect of thermo-meteorological parameters over the wheat yield was significant.  $T_{min}$  of November, December, March; GDD of November, December, March and the whole season were found to have significant negative correlation with wheat yield over wide spread districts of Zone (1) Similarly,  $T_{min}$  of November, December, January, March; GDD of November, December, March and the whole season showed significant negative correlation with wheat yield over majority of the districts of Zone (2)  $T_{min}$  of December, January, March; GDD of November, December, January, March and the whole season showed significant negative correlation with wheat yield over Zone (3)  $T_{max}$  did not show any consistent result across the zones. Important to note that  $T_{min}$  and its coupled parameter GDD during the latter part of the wheat season (ripening stage) plays a significant role in the wheat yield. Any increase in the thermal environment particularly in the later part of the wheat crop season has significant negative effect on the yield. It is apt to mention here that such scenarios of terminal heat stress on wheat crop are frequent over India in climate change scenarios. Hence, result of the present study could further assist in avoiding/ mitigating such stress towards sustainable wheat yield.

Overall, a region facing water scarcity such as Zone-01, wheat yield exhibits significant positive correlation with both seasonal and November month rainy days. The SPEI at a 3-month time scale, which captures the integrated response of rainfall and temperature, displays a significant positive correlation across all zones, especially during reproductive growth stage. Additionally, temperature-related parameters, including climatic extremes, show significant correlations. Specifically, an increase in minimum temperature adversely affects wheat yield in all zones. GDD, whether calculated seasonally or monthly, are inversely related to wheat yield.

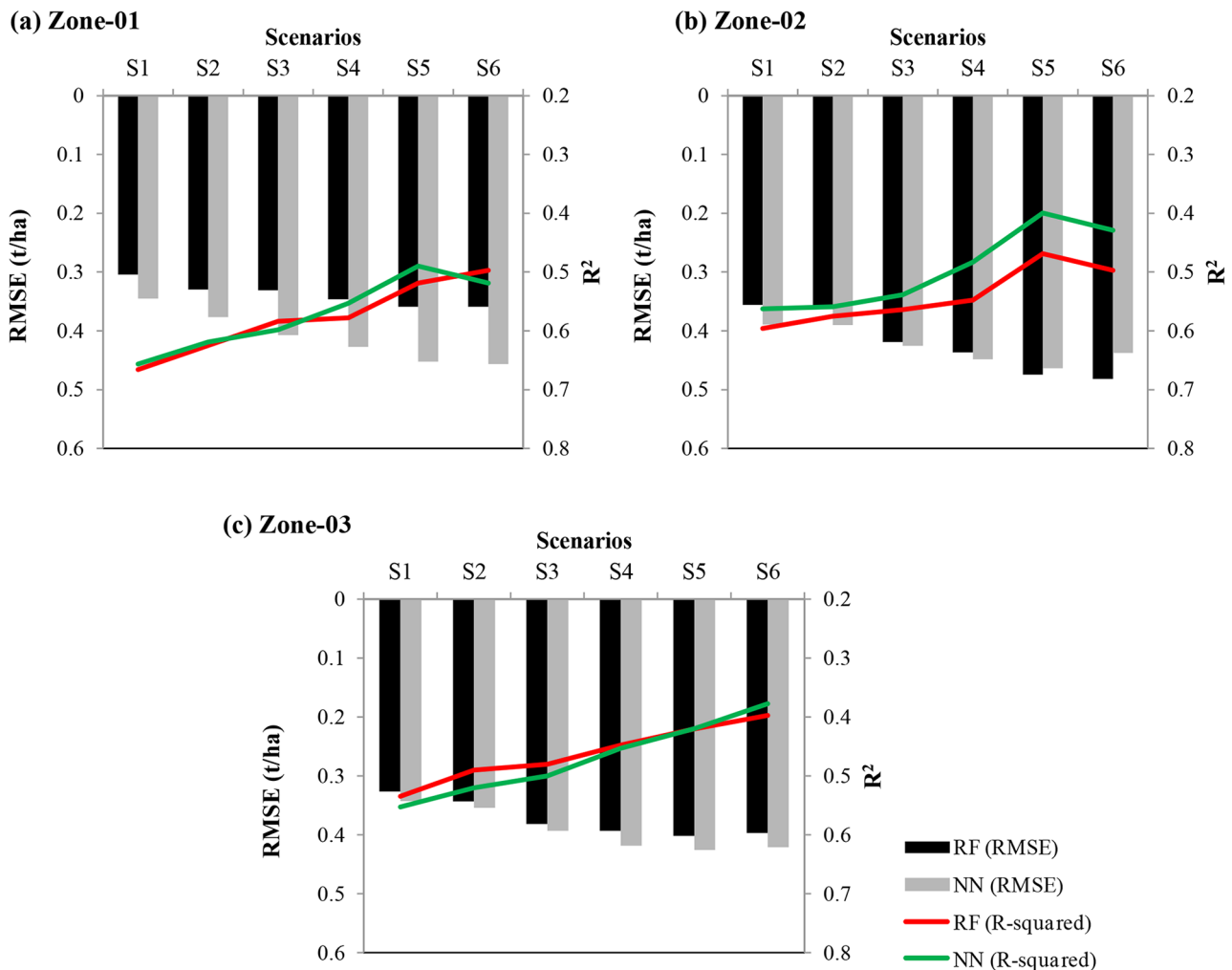
Moreover, the study identifies weather extremes such as a rise in the frequency of cold days and cold nights (indicating a decrease in maximum and minimum temperatures, respectively), positively affecting yield. Conversely, an increase in the occurrence of hot days and hot nights during the ripening phase negatively impacts wheat yield, aligning with previous studies (Rao et al. 2015; Dubey et al. 2020; Farhad et al. 2023). The observed yield reduction due to an increase in temperature may be related to increased crop water demand in heat-induced limited water conditions (Zhao et al. 2017; Zaveri and Lobell 2019). Furthermore, the study reveals a distinct effect of climate and its extremes varying with crop growth stages. In wheat, crown root initiation, which usually occurs after 21 days of sowing, is one of the most critical stages for water stress. Therefore, an increase

in temperature (Tmax) or a decrease in water availability (RD or SPEI) in November shows a negative association with crop yield. Other critical growth stages such as flowering, jointing, and milking stages corresponding to January, February, and March, respectively, also exhibit sensitivity towards temperature-related parameters (Tmax, frequency of hot days, cold days, and cold nights). Wheat yields in Madhya Pradesh exhibited a weak correlation with rainfall parameters across all zones. However, the Standardized Precipitation Evapotranspiration Index (SPEI) displayed a reasonable positive relationship with wheat yields. This suggests that SPEI serves as a better indicator of water availability for crops compared to rainfall, as it captures water stress effectively (Tirivarombo et al. 2018; Ortiz-Bobera et al. 2019).

## 4.2 Influence of predictive variables in crop yield estimation

### 4.2.1 Sensitivity analysis based on RMSE

Towards developing the impact of climate parameters on crop yield, sensitivity of causative parameters was assessed by systematically removing individual predictor variables in machine learning models. Six sets of scenarios were created as mentioned in Sect. 3.6. In the case of soybean, Scenario 1 (where all parameters were employed for prediction) exhibited the lowest RMSE in both RFR and NN models (0.3 t/ha and 0.34 t/ha in Zone-01; 0.35 t/ha and 0.39 t/ha in Zone-02; and 0.27 t/ha and 0.29 t/ha in Zone-03), as given in Fig. 3. To account for zone-wise variations in absolute yield, nRMSE was calculated, reflecting the residual variance in models and relating RMSE to the observed range of variables. Models employing all parameters also



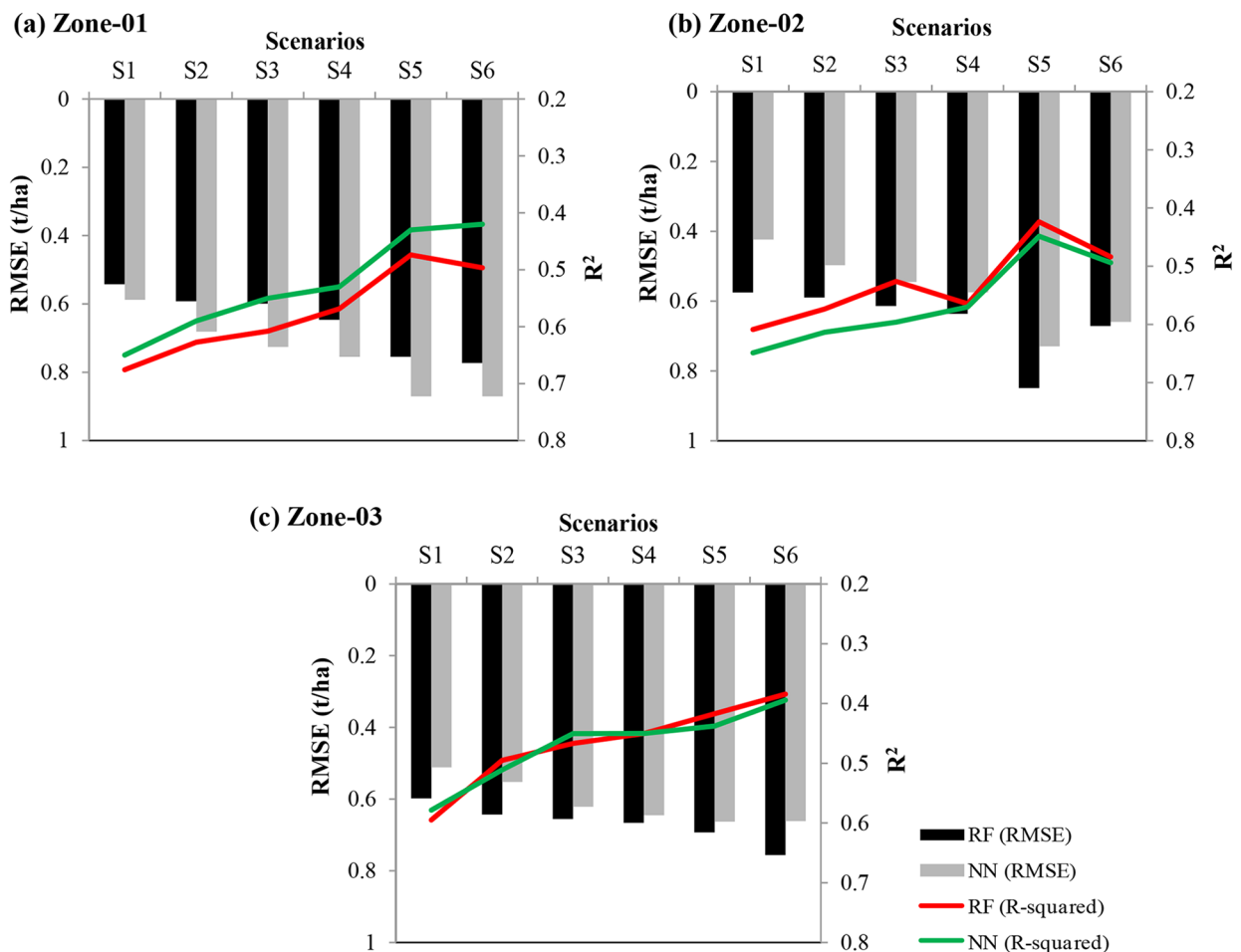
**Fig. 3** Error (RMSE) estimation and explained variance ( $R^2$ ) using various combinations of predictor variables for Soybean yield across different zones. *S1*: Scenario 1, *S2*: Scenario 2, *S3*: Scenario 3, *S4*: Scenario, *S5*: Scenario 5, *S6*: Scenario 6



yielded lowest normalized RMSE (nRMSE) values of 19% and 21% in Zone-01; 22.2% and 24.32% in Zone-02; and 15.2% and 16.2% in Zone-03 for RFR and NN models, respectively. Subsequent removal of individual parameters led to increases in both RMSE and nRMSE. Similarly, the  $R^2$  for both models in Scenario 1, encompassing all climatic variables and temperature extremes, could explain the variability in soybean yield to the extent of 65–66% in Zone-01, 56–59% in Zone-02, and 53–55% in Zone-03 (Fig. 3). Interestingly, the step-wise removals of thermal parameters such as temperature extremes,  $T_{max}$ ,  $T_{min}$ , and GDD as mentioned in Scenarios 2, 3, and 4 did not significantly impact the uncertainty estimates, suggesting that these parameters are intertwined in affecting the crop yield. A considerable portion of the influence of other temperature parameters on yields had already been captured by temperature extremes. For instance, an increase in the frequency of hot days inherently raises the monthly average  $T_{max}$ ,  $T_{min}$ , and heat accumulation (via GDD) during grain formation stages (Vogel et al. 2019). This intricate relationship indicates that

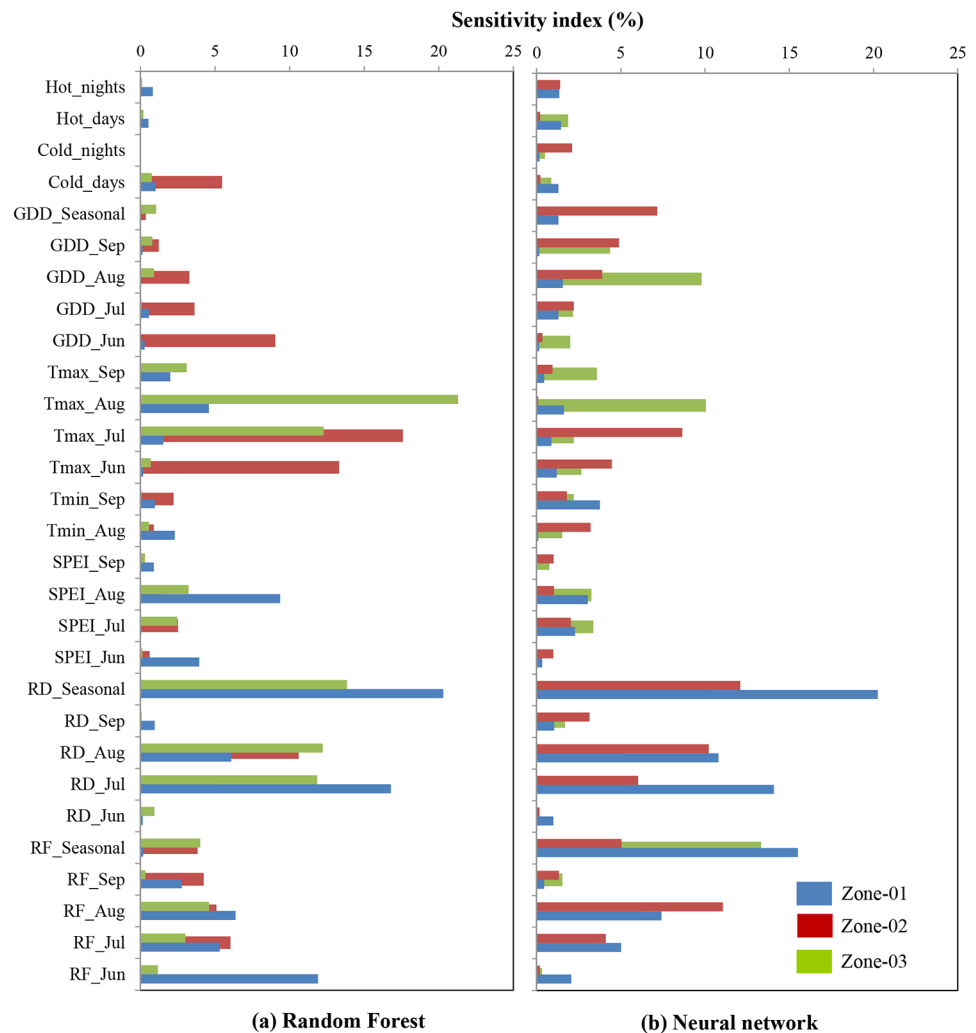
segregating the effect of temperature and its extremes on crop yield is challenging as these are often collinear. The maximum reduction (6–7%) in the explanatory power of the model was observed after the exclusion of hydro-meteorological parameters in Zone-01, whereas a decrease of 2–4% in explanatory power occurred after removing thermo-meteorological parameters in Zone-03. Notably, using only the rainfall parameter or SPEI parameter in RFR and NN models (RMSE 0.36 t/ha and 0.45 t/ha) increased the soybean yield estimation error by 18% and 32%, respectively, compared to when all parameters were utilized.

Similarly, all predictor variables (Scenario 1) could able to estimate wheat yield and explain the variability with RMSE (and  $R^2$ ) values of 0.54 t/ha (0.67) and 0.58 t/ha (0.65) in Zone-01; 0.57 t/ha (0.60) and 0.42 t/ha (0.64) in Zone-02; and 0.59 t/ha (0.59) and 0.51 t/ha (0.57) in Zone-03 using RFR and NN, respectively (Fig. 4). RFR and NN models utilizing all parameters exhibited nRMSE values of 21.7% and 23.5% (Zone-01); 25.5% and 18.8% (Zone-02); and 20% and 17% (Zone-03), respectively. The stepwise



**Fig. 4** Error (RMSE) estimation and explained variance ( $R^2$ ) using various combinations of predictor variables for Wheat yield across different zones. **S1**: Scenario 1, **S2**: Scenario 2, **S3**: Scenario 3, **S4**: Scenario, **S5**: Scenario 5, **S6**: Scenario 6

**Fig. 5** Sobols' sensitivity index of different predictor variables of Soybean yield over different zones



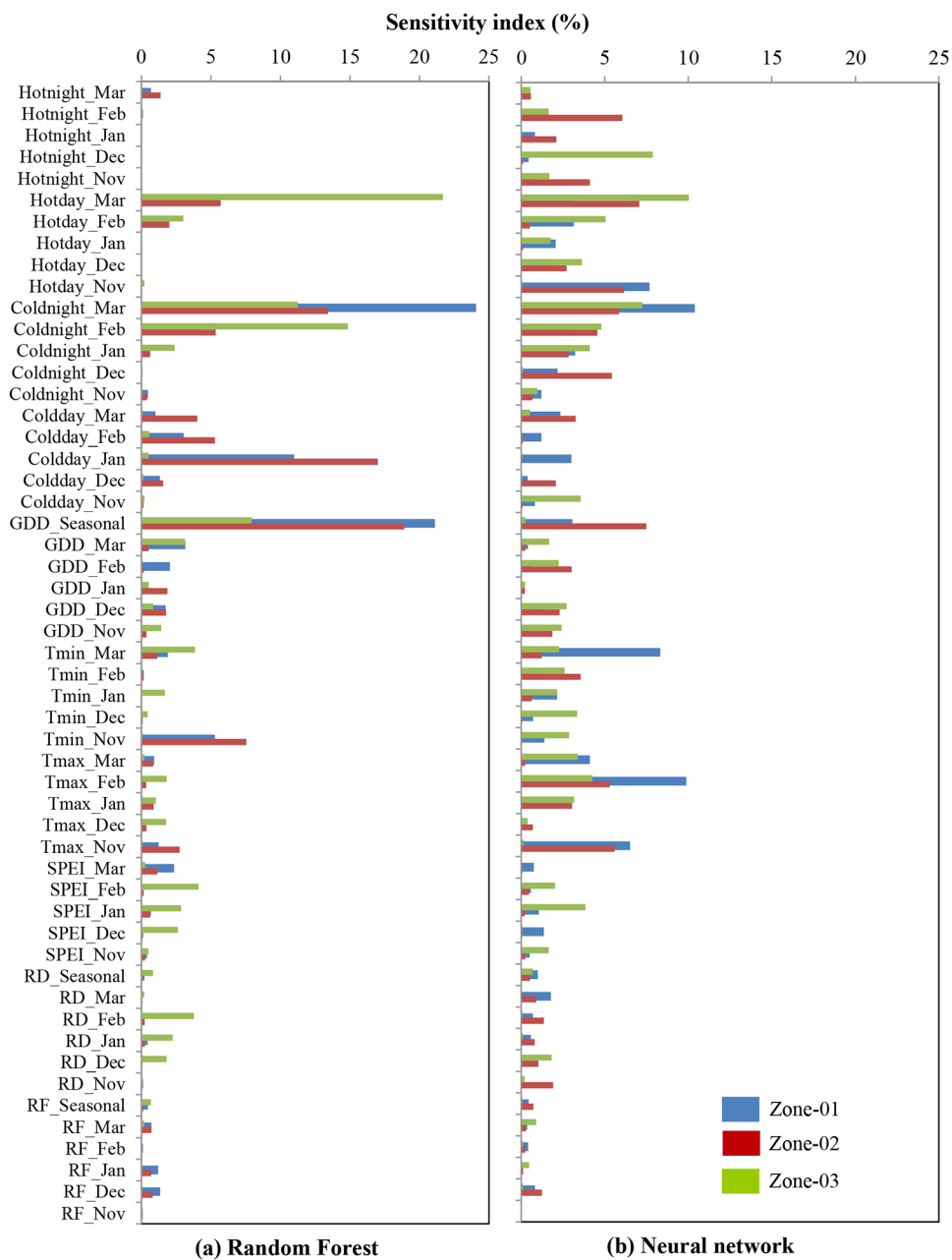
removal of predictor variables increased both RMSE and nRMSE. In Zone-01, models incorporating only hydro-meteorological parameters and excluding thermal parameters (Scenario 4) resulted in a 20% increase in RMSE and 10% reduction in model explanatory capability. Conversely, models working with either RF/RD or SPEI (Scenarios 5 or 6) showed increased RMSE (and decreased  $R^2$ ) and therefore, accounted the less variability in crop yield, underscoring their significant role in wheat yield in Zone-01 with low rainfall. While increases in RMSE were observed in Zone-03, the extent was not as high as water limited Zone-01 and Zone-02, indicating that thermal parameters explained most of the variability. It is worth to note that both state-of-the-art ML models captured the similar pattern in all scenarios despite of different extent of RMSE values.

#### 4.2.2 Sensitivity analysis of different predictor variables using Sobol' index

Sobol sensitivity analysis is commonly employed for complex system models to quantitatively decompose output variance concerning input parameters. In this study, a first-order Sobol' index was utilized to assess the individual impact of input parameters on the model while keeping all other parameters fixed. The first-order Sobol' index always yields positive values (direction-less), and the sum of all individual parameters should be equal to 1. Parameters with sensitivity indices greater than 5% are deemed significant (Zhang et al. 2015).

For soybean yield, the study revealed sensitivity to seasonal RD (>20%), followed by RD in July and August in Zone-01 (Fig. 5). This underscores water stress as a limiting factor for yield during the crop development stages (from leaf and stem development to flowering) in regions facing water scarcity. In Zone-02, characterized by high rainfall but poor soil water retention, Tmax plays a major role in

**Fig. 6** Sobols’ sensitivity index of different predictor variables of Wheat yield over different zones



soybean yield variability in both Random Forest (RFR) and Neural Network (NN) models. Tmax accelerates leaf senescence and restricts photosynthetic activity, consequently affecting soybean pod filling. In Zone-03, both temperature and rainfall during soybean’s reproductive growth stages (flowering to pod filling) are critical factors influencing yield.

In the case of wheat crops, the study found sensitivity to spells of cold nights (i.e., frequency of Tmin lower than the 10th percentile) during the reproductive phase in both Zone-01 and Zone-02. Other parameters influencing wheat yield include thermal factors such as GDD, Tmax, cold days, and

hot days. In Zone-03, temperature extremes like spells of hot days and cold nights in March emerged as crucial factors affecting wheat yield (Fig. 6). It is a well-known fact that exposure of wheat to extreme heat events during the reproductive phase leads to sterile flowers, damaged pollen tube growth, and fertilization issues, resulting in reduced wheat yield (Shenoda et al. 2001; Ullah et al. 2022). Wheat yields in the central part of India appear unresponsive to hydro-meteorological parameters (<5% Sobol’ index) across all growth stages in all zones, aligning with findings from similar studies conducted elsewhere (Birthal et al. 2014; Petersen 2019; Schierhorn et al. 2021). Notably, an agreement exists

between RFR and NN models in terms of the relative importance of climatic factors affecting soybean and wheat crop yield, despite differences in Sobol' index magnitude.

## 5 Discussion

While previous studies have examined the influence of in-season climatic variables on crop yield (Harkness et al. 2020), there has been relatively little research linking these variables to specific crop development phases and their contribution to understanding the relationship between weather and crop yield using machine learning (ML) techniques (Hofman et al. 2020; Beillouin et al. 2020; Schierhorn et al. 2021). Assuming zone-wise consistent soil conditions, we focus on climate variability as the primary determinant of soybean and wheat crop yield in each zone. Our results indicate that climatic variables can explain a significant proportion of yield variability, ranging from 66 to 42% in zone-01 and 59–37% in zone-03, regardless of crop type. Additionally, weather extremes play a crucial role during specific crop growth stages, although their impact is less pronounced compared to climatic variables, as observed by Schierhorn et al. (2021). These climate-driven impacts vary not only with crop types such as soybean or wheat but also with the prevailing climate and soil characteristics of the region (Kukul and Irmak 2018). In our study, we found that the effects of heat and water stress vary significantly among different crop types and regions. Vicente-Serrano et al. (2013) noted reduced vegetation sensitivity to water stress in regions with climatological water excess, consistent with our findings. For instance, hydro-meteorological parameters were not found to be sensitive in Zone-03, which does not lack rainfall. Moreover, wheat, even as an irrigated crop, did not exhibit sensitivity to hydro-meteorological parameters leading to yield decline.

Murari et al. (2018) emphasized the significance of temperature extremes on crop yield, particularly in the southern part of India. Hofman et al. (2020) also found a positive response of soybean to increases in cool minimum temperatures using random forest (RF) models. In our study, we did not find a significant association between cold nights (minimum temperature extremes) and soybean yield in Indian conditions. However, a decrease in soybean yield was observed with high maximum temperatures (Tmax) during the pod filling stage, consistent with findings for soybean yield sensitivity to August Tmax in Zone-03. Sharma et al. (2022) reported a 25.6% decline in soybean yield with a 1 °C rise in Tmax. In wheat crops, extreme heat during the grain-filling phase, particularly in Zone-03, reduced yield, as heat stress accelerates grain filling, reducing both duration and ultimately grain yield (Barlow et al. 2015). Hot

nights during February negatively affected wheat yields during the anthesis phase, reducing photosynthesis and increasing respiration, leading to yield loss (Sadok and Jagadish 2020). Rao et al. (2015) found a significant negative relation between wheat yield and increased Tmax and Tmin in January and February in India, with Tmin exhibiting a more pronounced impact. Various studies have reported different degrees of yield decline under temperature effects, depending on the location of the study. For instance, Rao et al. (2015) documented a 7% yield loss with a 1 °C rise in Tmin in India, while You et al. (2009) reported a range of 3–10% in China. Additionally, Lobell et al. (2005) observed a 10% decline in Mexico, among others. It was observed that Tmin exceeding 12 °C and Tmax exceeding 34 °C were identified as critical during the post-anthesis period (Rao et al. 2015). Our study indicates that in zones with limited moisture and soil water-holding capacity, Tmin extremes and seasonal GDD are determining factors for wheat yield. However, in Zone-03 (without limitations), extreme Tmax in March, representing the grain filling and ripening period, emerges as a sensitive parameter for wheat yield. Additionally, cold nights (Tmin extremes) during the reproductive phase of wheat were identified as one of the most sensitive parameters, along with seasonal GDD, using global sensitivity analysis.

The outcome of present study may align to several adaptation strategies useful to mitigate the impact of climate and weather extremes on wheat and soybean crop yield, including altering cropping patterns, practicing minimum tillage (Biswas et al. 2008), adjusting sowing dates, cultivating heat-tolerant varieties, and implementing additional fertilizer and irrigation (Dubey et al. 2020).

## 6 Limitations and future scope

Our study emphasizes that utilizing intra-seasonal climatic variables for distinct crop developmental stages enhances the understanding of the relationship between weather and yield during these critical periods, aligning with previous research findings (Schierhorn et al. 2021). A deeper comprehension of intra-seasonal climate effects on yields can provide valuable insights for decision-making processes among farmers and policymakers. However, it is noteworthy that the current study has not utilized the explicit consideration of phenology. Instead, it employs monthly weather and climatic parameters as representatives of the dominant growth stage occurring in that respective month. As a future scope of work, we propose for a more distinct approach by integrating satellite-derived crop growth stages, encompassing key phases like the start of the season, peak growth stage, reproductive stage, and senescence, which will enhance the



precision of the analysis. Moreover, the current study has utilized district-level yield data, which sometimes nullifies the impact of spatial variability of soil and rainfall on crop yield. Therefore, the utilization of sub-district level yield data holds promises for providing more granular insights to capture the localized effects of weather and climate dynamics more effectively. To comprehensively enrich the model, it is imperative to extend its scope by incorporating additional meteorological parameters, specifically addressing extreme events such as rainfall extremes, hailstorms, floods, among others. This inclusion will undoubtedly contribute to a more holistic and robust analysis, equipping farmers and policymakers with more informed decision-making tools to face the climate vagaries.

## 7 Conclusions

Climate variability significantly influences crop yield during both the monsoon and winter seasons, as demonstrated by studies on soybean and wheat crops. The impact of climatic variables and extremes varies in intensity and extent based on the biophysical characteristics of the region. To explore this phenomenon, the present study was conducted across homogenous zones categorized by rainfall patterns and soil parameters. Monthly and seasonal climatic variables and extremes were evaluated to assess their impact on the critical growth stages of the crops. Machine Learning (ML) techniques such as, RFR and NN were employed to understand the complex relationship between climate/weather extremes vis-a vis crop yield. RF successfully predicted soybean and wheat yields with RMSE values of 0.28–0.38 t/ha and 0.45–0.6 t/ha, respectively. Similarly, NN predicted yields with RMSE values of 0.27–0.39 t/ha and 0.4–0.58 t/ha for soybean and wheat, respectively. These findings highlight the accuracy of the models in capturing the intricate relationship between climate variables and crop yield.

The study revealed that rainfall and its derivatives significantly influence soybean crop yield during the monsoon season. In contrast, temperature and its extremes dictate wheat crop yield in the winter season. However, these effects exhibit variations across different zones within a season. In regions with limited water supply, parameters such as rainfall and rainy-days in July–August play a pivotal role in determining soybean crop yield, corresponding to the flowering and pod-filling stages. Conversely, in well-irrigated regions, temperature-related parameters exert control over crop yield. Cold nights and hot days during reproductive growth stages demonstrated sensitivity concerning wheat yield, with positive and negative relationships, respectively. Considering the anticipated increase in climatic extremes in the future, gaining a deeper understanding of the intricate

relationship between climate and yield variability is imperative. This knowledge will pave the way for enhanced adaptation strategies, mitigating potential yield losses and ensuring sustainable agricultural practices.

**Acknowledgements** Authors thank Director, National Remote Sensing Centre (NRSC), Hyderabad, India and Deputy Director, Remote Sensing Applications, NRSC for providing facility to carry out the work. We thank Dr. VM Chowdary, Group Director, Agricultural Sciences and Applications Group, NRSC for his valuable suggestions. Authors also acknowledge the support provided by University of Hyderabad (UoH), Hyderabad, India. Authors thank to the reviewers for their valuable feedback and insightful suggestions.

**Author contributions** M.K. and A.C. conceptualized the work, framed methodology, and wrote the main manuscript text. M.K. curated data and carried out analysis. A.C., V.C. and P.S.R. supervised the work and reviewed the manuscript. V.P. contributed to the formal analysis.

**Funding** The authors declare that no funds and grants were received during the preparation of this manuscript.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Competing interests** The authors declare no competing interests.

**Conflict of interest** The authors declare no conflict of interest.

## References

- Amit Y, Geman D (1997) Shape quantization and recognition with randomized trees. *Neural Comput* 10(7):1545–1588
- Barlow KM, Christy BP, O’Leary GJ, Riffkin PA, Nuttall JG (2015) Simulating the impact of extreme heat and frost events on wheat crop production: a review. *Field Crops Res* 171:109–119
- Beillouin D, Schauburger B, Bastos A, Ciais P, Makowski D (2020) Impact of extreme weather conditions on European crop production in 2018. *Philos Trans R Soc B: Biol Sci* 375:20190510
- Birthal PS, Khan T, Negi DS, Agarwal S (2014) Impact of climate change on yields of major food crops in India: implications for food security. *Agricultural Econ Res Rev* 27(2):145–155
- Biswas DK, Xu H, Li YG, Liu MZ, Chen YH, Sun JZ, Jiang GM (2008) Assessing the genetic relatedness of higher ozone sensitivity of modern wheat to its wild and cultivated progenitors/relatives. *J Exp Bot* 59(4):951–963
- Braun HJ, Rajaram S, Ginkel M (1997) CIMMYT’s approach to breeding for wide adaptation. In *Adaptation in plant breeding: selected papers from the XIV EUCARPIA Congress on Adaptation in Plant breeding held at Jyväskylä, Sweden from July 31 to August 4, 1995* (pp. 197–205). Springer Netherlands
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Chakraborty A, Seshasai MV, Rao SK, Dadhwal VK (2017) Geo-spatial analysis of temporal trends of temperature and its extremes over India using daily gridded (1×1) temperature data of 1969–2005. *Theoretical Appl Climatol* 130:133–149
- Chakraborty D, Sehgal VK, Dhakar R, Ray M, Das DK (2019) Spatio-temporal trend in heat waves over India and its impact assessment on wheat crop. *Theoret Appl Climatol* 138:1925–1937

- Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ Res Lett* 13(11):114003
- de Wit CT (1965) Photosynthesis of leaf canopies. Pudoc; 1965
- Dubey R, Pathak H, Chakrabarti B, Singh S, Gupta DK, Harit RC (2020) Impact of terminal heat stress on wheat yield in India and options for adaptation. *Agric Syst* 181:102826
- Farhad M, Kumar U, Tomar V, Hossain A (2023) Heat stress in wheat: a global challenge to feed billions in the current era of the changing climate. *Front Sustainable Food Syst* 7:1203721
- Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics 31 (pp. 249–256). JMLR Workshop and Conference Proceedings
- Gupta S, Bhandari L, Jakhu R, Sharma M (2022) Climate change, weather anomalies, and agriculture: impact on output of major crops in India (CSEP Working Paper 46). New Delhi: Centre for Social and Economic Progress
- Harkness C, Semenov MA, Areal F, Senapati N, Trnka M, Balek J, Bishop J (2020) Adverse weather conditions for UK wheat production under climate change. *Agric for Meteorol* 282:107862
- Harvey C (2018) Scientists can now blame individual natural disasters on climate change extreme event attribution is one of the most rapidly expanding areas of climate science. *Scientific American*
- Hofman A, Kemanian A, Forest C (2020) The response of maize, sorghum, and soybean yield to growing phase climate revealed with machine learning. *Environ Res Lett* 15:094013
- IPCC, Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M (2021) Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change. 2021;2
- Kang Y, Khan S, Ma X (2009) Climate change impacts on crop yield, crop water productivity and food security—a review. *Prog Nat Sci* 10(1912):1665–1674
- Konduri VS, Vandal TJ, Ganguly S, Ganguly AR (2020) Data science for weather impacts on crop yield. *Front Sustainable Food Syst* 19:4:52
- Kukul MS, Irmak S (2018) Climate-driven crop yield and yield variability and climate change impacts on the US Great Plains agricultural production. *Sci Rep* 22(1):1–8
- Liaw A, Wiener M (2002) Classification and regression by random Forest. *R news*. 3;2(3):18–22
- Lobell DB, Asseng S (2017) Comparing estimates of climate change impacts from process-based and statistical crop models. *Environ Res Lett* 4(1):015001
- Lobell DB, Ortiz-Monasterio JI, Asner GP, Matson PA, Naylor RL, Falcon WP (2005) Analysis of wheat yield and climatic trends in Mexico. *Field Crops Res* 94(2–3):250–256
- Lobell DB, Sibley A, Ivan Ortiz-Monasterio J (2012) Extreme heat effects on wheat senescence in India. *Nat Clim Change* 2(3):186–189
- Lu J, Carbone GJ, Gao P (2017) Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. *Agric for Meteorol* 1:237:196–208
- MacCarthy DS, Traore PS, Freduah BS, Adiku SG, Dodor DE, Kumhar SK (2022) Productivity of soybean under projected climate change in a Semi-arid region of West Africa: sensitivity of current production system. *Agronomy* 24(11):2614
- Madhukar A, Dashora K, Kumar V (2021) Climate trends in temperature and water variables during wheat growing season and impact on yield. *Environ Processes* 8:1047–1072
- Matiu, Matiu M, Ankerst DP, Menzel A et al (2017) 2017. Interactions between temperature and drought in global and regional crop yield variability during 1961–2014. *PloS one*. 26;12(5):e0178339
- Mohanty M, Sinha NK, McDermid SP, Chaudhary RS, Reddy KS, Hati KM, Somasundaram J, Lenka S, Patidar RK, Prabhakar M, Cherukumalli SR (2017) Climate change impacts vis-a-vis productivity of soybean in vertisol of Madhya Pradesh. *J Agrometeorology* 1(1):10–16
- Mohapatra S, Mohapatra S, Han H, Ariza-Montes A, López-Martín MD (2022) Climate change and vulnerability of agribusiness: assessment of climate change impact on agricultural productivity. *Front Psychol* 26:13:955622
- Moore FC, Baldos UL, Hertel T (2017) Economic impacts of climate change on agriculture: a comparison of process-based and statistical yield models. *Environ Res Lett* 12(6):065008
- Murari KK, Mahato S, Jayaraman T, Swaminathan M (2018) Extreme temperatures and crop yields in Karnataka, India. *Rev Agrarian Studies*. 2018;8(2)
- Novikova LY, Bulakh PP, Nekrasov AY, Seferova IV (2020) Soybean response to weather and climate conditions in the Krasnodar and Primorye territories of Russia over the past decades. *Agronomy* 10(9):1278
- Ortiz-Bobea A, Wang H, Carrillo CM, Ault TR (2019) Unpacking the climatic drivers of US agricultural yields. *Environ Res Lett* 14:064003
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32
- Paymard P, Bannayan M, Haghighi RS (2018) Analysis of the climate change effect on wheat production systems and investigate the potential of management strategies. *Nat Hazards* 91:1237–1255
- Peichl M, Thober S, Meyer V, Samaniego L (2018) The effect of soil moisture anomalies on maize yield in Germany. *Nat Hazards Earth Syst Sci* 18:889–906
- Petersen LK (2019) Impact of climate change on twenty-first century crop yields in the U.S. *Climate* 7:40
- Porter JR, Gawith M (1999) Temperatures and the growth and development of wheat: a review. *Eur J Agron* 10(1):23–36
- Powell JP, Reinhard S (2016) Measuring the effects of extreme weather events on yields. *Weather Clim Extremes* 1:12:69–79
- Praveen D, Palanivelu K (2017) Spatiotemporal analysis of projected impacts of climate change on the major C3 and C4 crop yield under representative concentration pathway 4.5: insight from the coasts of Tamil Nadu, South India. *PLoS ONE* 12(7):e0180706
- Rao BB, Chowdary PS, Sandeep VM, Pramod VP, Rao VU (2015) Spatial analysis of the sensitivity of wheat yields to temperature in India. *Agric for Meteorol* 15:200:192–202
- Robert M, Thomas A, Sekhar M, Badiger S, Ruiz L, Raynal H, Bergez JE (2017) Adaptive and dynamic decision-making processes: a conceptual model of production systems on Indian farms. *Agricultural Syst* 1:157:279–91
- Sadok W, Jagadish SK (2020) The hidden costs of nighttime warming on yields. *Trends Plant Sci* 25(7):644–651
- Schierhorn F, Hofmann M, Gagalyuk T, Ostapchuk I, Müller D (2021) Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages. *Clim Change* 169(3–4):39
- Sharma RK, Kumar S, Vatta K, Dhillion J, Reddy KN (2022) Impact of recent climate change on cotton and soybean yields in the south-eastern United States. *J Agric Food Res* 9:100348
- Shenoda JE, Sanad MN, Rizkalla AA, El-Assal S, Ali RT, Hussein MH (2021) Effect of long-term heat stress on grain yield, pollen grain viability and germinability in bread wheat (*Triticum aestivum* L.) under field conditions. *Heliyon*, 7(6).
- Sobol WT (1993) Analysis of variance for ‘component stripping’ decomposition of multiexponential curves. *Comput methods programs in biomed*, 39(3–4), 243–257

- Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul* 55(1–3):271–280
- Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. *BMC Bioinformatics* 9:1–1
- Subramaniam AR (1983) Agro-ecological zones of India. *Archives for meteorology, geophysics, and bioclimatology. Ser B* 32(2):329–333
- Tigchelaar M, Battisti DS, Naylor RL, Ray DK (2018) Future warming increases probability of globally synchronized maize production shocks. *Proc Nat Acad Sci* 26;115(26):6644–9
- Tirivarombo S, Osupile D, Eliasson P (2018) Drought monitoring and analysis: standardised precipitation evapotranspiration index (SPEI) and standardised precipitation index (SPI). *Phys Chem Earth Parts A/B/C* 106(1):1–0
- Ullah A, Nadeem F, Nawaz A, Siddique KH, Farooq M (2022) Heat stress effects on the reproductive physiology and yield of wheat. *J Agron Crop Sci* 208(1):1–7
- Van Oort PA, Zhang T, De Vries ME, Heinemann AB, Meinke H (2011) Correlation between temperature and phenology prediction error in rice (*Oryza sativa* L). *Agric for Meteorol* 15(12):1545–1555
- Vicente-Serrano SM, Gouveia C, Camarero JJ, Beguería S, Trigo R, López-Moreno JJ, Azorín-Molina C, Pasho E, Lorenzo-Lacruz J, Revuelto J, Morán-Tejeda E (2013) Response of vegetation to drought time-scales across global land biomes. *Proc Nat Acad Sci* 110(1):52–7
- Vogel E, Donat MG, Alexander LV, Meinshausen M, Ray DK, Karoly D, Meinshausen N, Frieler K (2019) The effects of climate extremes on global agricultural yields. *Environ Res Lett* 14:054010
- Webber H, Ewert F, Olesen JE, Müller C, Fronzek S, Ruane AC, Bourgault M, Martre P, Ababaei B, Bindi M, Ferrise R, Finger R, Fodor N, Gabaldón-Leal C, Gaiser T, Jabloun M, Kersebaum K-C, Lizaso JJ, Lorite IJ, Manceau L, Moriondo M, Nendel C, Rodríguez A, Ruiz-Ramos M, Semenov MA, Siebert S, Stella T, Stratonovitch P, Trombi G, Wallach D (2018) Diverging importance of drought stress for maize and winter wheat in Europe. *Nat Commun* 9:4249
- Yu Y, Huang Y, Zhang W (2013) Projected changes in soil organic carbon stocks of China's croplands under different agricultural managements, 2011–2050. *Agric Ecosyst Environ* 15(178):109–120
- Zampieri M, Ceglar A, Dentener F, Toreti A (2017) Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environ Res Lett* 12:064008
- Zaveri E, Lobell D (2019) The role of irrigation in changing wheat yields and heat sensitivity in India. *Nat Commun* 10(1):9
- Zhang XY, Trame MN, Lesko LJ, Schmidt S (2015) Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. *CPT: Pharmacometrics Syst Pharmacol* 4(2):69–79
- Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciaia P, Durand JL, Elliott J, Ewert F, Janssens IA, Li T, Lin E, Liu Q, Martre P, Müller C, Peng S, Peñuelas J, Ruane AC, Wallach D, Wang T, Wu D, Liu Z, Zhu Y, Zhu Z, Asseng S (2017) Temperature increase reduces global yields of major crops in four independent estimates. *Proc Nat Acad Sci* 114(35):9326–9331

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.