

# A Novel Borda Count Based Feature Ranking and Feature Fusion Strategy to Attain Effective Climatic Features for Rice Yield Prediction

Subhadra Mishra

Department of Computer Science and Application, CPGS  
Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India  
E-mail: mishra.subhadra@gmail.com

Debahuti Mishra

Department of Computer Science and Engineering,  
Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha, India  
E-mail: debahutimishra@soa.ac.in

Pradeep Kumar Mallick

School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India  
E-mail: pradeep.mallickfcs@kiit.ac.in

Gour Hari Santra

Department of Soil Science and Agricultural Chemistry, IAS  
Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha, India  
E-mail: santragh@gmail.com

Sachin Kumar (Corresponding Author)

Department of Computer Science, South Ural State University, Chelyabinsk, Russia  
E-mail: sachinagnihotri16@gmail.com

**Keywords:** rice crop yield prediction, climatic variability, extreme learning machine, feature ranking, feature fusion

**Received:** July 29, 2020

*An attempt has been made in the agricultural field to predict the effect of climatic variability based on rice crop production and climatic features of three coastal regions of Odisha, a state of India. The novelty of this work is Borda Count based fusion strategy on the ranked features obtained from various ranking methodologies. Proposed prediction model works in three phases; in the first phase, three feature ranking approaches such as; Random Forest, Support Vector Regression-Recursive Feature Elimination (SVR-RFE) and F-Test are applied individually on the two datasets of three coastal areas and features are ranked as per their algorithm. In the second phase; Borda Count as a fusion method has been implemented on those ranked features from the above phase to obtain top five best features. The multiquadratic activation function based Extreme Learning Machine (ELM) has been used to predict the rice crop yield using those ranked features obtained from fusion based ranking strategy and the number of varying features are obtained which gives prediction accuracy above 99% in the third phase of experimentation. Finally, the statistical paired T-test has been used to evaluate and validate the significance of proposed fusion based ranking prediction model. This prediction model not only predicts the rice yield per hectare but also able to obtain the significant or most affecting features during Rabi and Kharif seasons. From the observations made during experimentation, it has been found that; relative humidity is playing a vital role along with minimum and maximum temperature for rice crop yield during Rabi and Kharif seasons.*

*Povzetek: Članek opisuje izviren pristop pri iskanju vzorcev vremenske variabilnosti s pomočjo metod za izbiro in združevanjem atributov.*

## 1 Introduction

Agriculture is the major source of livelihood for people in Odisha as well as India, but here it is said that 'Agriculture is the gamble of the monsoon'. Due to the climatic changes the production of major yield is reduced in the Kharif. While Kharif rain fall over the country might be increased by 10-15%, but winter rain fall is expected to de-

crease by 5-25% and seasonal variability would be further compounded [1].

It is highlighted that, due to heavy temperature, including water shortage, distribution of rainy days, maximum loss is expected in Rabi crops and the productivity of Rabi crops is decreased from 10% to 40% by 2100 [2]. Rice yield is expected to decline by 6% for every 10°C rise in

temperature [3]. The scientific and policy personnel have accepted the susceptibility of agriculture crop to climate change and raised question the capability of farmers to adapt because of the direct and strong dependence of crop agriculture on climate [4]. There are different forecasting methodologies available and evaluated by the research workers all over the world in the field of Agriculture. On all India basis, the imitation study developed shows that the yield of rice crop is affected by weather change from 2.5 to 12% [5]. The rice is the main food in eastern India specifically in the states of Odisha, West Bengal, Jharkhand and Bihar. In India green revolution is mainly Wheat as contributed states was mainly Punjab, Haryana and UP. So, Government of India is expecting the 2<sup>nd</sup> green revolution from eastern India. The amount of data set is very large in Indian agriculture. Earlier, the different model form dataset was done only by manual system, when there was no outset of computer. But with advancement of computer technology, collection of huge data, their classification and storage has been increased. This has established enormous improvement in pattern perception. In this paper, the main focus to develop a user friendly network for farmers which provide the study of rice production on the basis of important climatic parameter.

The current age is the age of data. As we are taking the large dataset for accuracy of the result, so for modeling of the dataset the feature selection technique becomes the prerequisite method [6, 7]. To increase the correctness level of the experiment we have to increase the attributes of the training examples that is the dataset [8, 9, 10]. As the knowledge discovery technique is finding the knowledge from the vast amount of data, so it is dare to do future research for solving the real world troubles. Ranking is a method to find a rank between all the features according to their importance. Selecting a least number of features produce a simple model, this will take less time for computation and can be understood easily. Due to the simpler model fewer resources also required, which can be affordable. Now the question is how we can rank the features or variables [9, 10, 11, 12, 13, 14]. There are so many algorithms in machine learning to find the significant variables. Thus, the concept of feature selection or variable selection arises. It is the selection of the variables or selecting the subset of the variables and this technique does not change the original illustration of the variables.

During the application of the various feature ranking techniques on the dataset, on each iteration small subsets are being generated. For each feature, there is a rank order of the result of each run and then united with the earlier runs to form an ensemble [15, 16]. The Monte Carlo algorithm states that an conclusion can be achieved by the combining random consecutive rough calculation to the same result [17]. This method stimulated the ensemble method.

As agriculture is the backbone bone of Indian economy and rice is the main staple food, so the prediction of rice and the timely advice on variation of climatic condition for the farmers is required. This factor motivate us to pre-

pare a computational model for the farmers and ultimately to the society also. The main aim of this work is to prepare a computational model to find the feature affected most for the rice production. Here we have used three different feature ranking methods such as Random Forest [18, 19, 20, 21, 22, 23], SVR-RFE [24, 25, 26] and F-Test [27, 28] for regression. These are mainly used for ranking of genes in gene expression datasets. The same methods are used here to rank the features of rice crop prediction datasets. Three ranking algorithms gave three different ranks to each feature of the dataset. Then, a feature fusion method has been proposed to evaluate the final rank of each feature and then, these newly ranked features are evaluated by Extreme Learning Machine (ELM) [29, 30, 31, 32] based regressor to measure the importance of each feature. The accuracy of ELM-Regressor has been calculated by decreasing one by one feature from the dataset. Finally, the comparison between proposed fusion based ranking strategy and non fusion based ranking strategy has been made to obtain the number significant features contributing towards the maximum accuracy of regressor. These features decide the importance of climatic parameters in rice crop production both for the Rabi season and Kharif season in the collected districts namely, Balasore, Cuttack and Puri. Thus the important finding of the study is temperature and humidity affect mostly for the crop production in the coastal district of Odisha.

## 1.1 Study area

In the Figure 1, the rice crop production dataset of three districts such as: Balasore, Puri, Cuttack are shown [33]. The production of rice is mainly in two seasons, such as: Rabi and Kharif. There are different features considered for this production, such as: rainfall, minimum and maximum temperature and relative humidity in the morning and afternoon hour. To avoid the inconsistency in the dataset there are various methods for missing value [36] imputation. In this paper mean value used to solve the missing value problem.

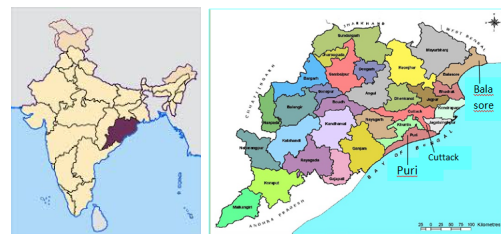


Figure 1: Odisha complete area taken from Google Map an state of India [34]

## 1.2 Goal

Considering the typical data available in the above mentioned section, the use of data mining or machine learning

strategies should be able to produce a natural decision for crop production based on the important or significant climatic parameters which affects the yield of rice during both the Rabi and Kharif seasons. This paper mainly focuses on the capabilities of ranking and fusion strategies, on two aspects such as; feature ranking and fusion of those ranked features. Specifically, the goal of this study can be outlined as follows:

- (a) Collection of climatic data of rice yield for both the Rabi season and the Kharif season of three coastal areas of Odisha, a state of India.
- (b) Feature importance evaluation and selection;
  - (i) Ranking of features by applying various ranking strategies.
  - (ii) Fusion of those ranked features.
- (c) Selection important climatic features derived from the ranked and fused features.
- (d) Model tuning or searching for appropriate algorithm parameters for better performance.
- (e) Model evaluation and validation through performance comparisons and statistical validation.

### 1.3 Paper layout

The rest of the paper is outlined as follows; the related work in this field is discussed in Section 2. The diagrammatic representation of proposed regressor has been detailed in Section 3. The methodologies such as Random Forest, SVR-RFE, F-Test and ELM regressor and various fusion strategies are discussed in Section 4. The experimentation and model evaluation is discussed in Section 5 and Section 6 discusses the principal findings obtained from this study. Finally Section 7 concludes the paper with future scope of this work.

## 2 Literature survey

To contextualize the effect of goals set and discussed in Section 1.2 in rice yield modeling, many papers were selected for review which are based on machine learning or data mining techniques be useful for modeling in this serial; (a) ranking of features based on Random Forest, F-Test and SVR-RFE (b) fusion strategies for feature selection and; (c) model evaluation and validation for proper classification. This section explores the various works done on prediction on agricultural field based on random forest, F-Test and SVR-RFE etc. SML Venkata et al. [35] used the dataset consisting of rainfall, precipitation and temperature and applied random forest which is the collection of decision trees, on the two-third of the records and then the resulting decision trees are applied on the remaining records and lastly for the prediction of the crop data, the resultant

training sets applied on the test data based on the input attributes. They have used R Studio and they evaluated their results by using other performance measures. Evathia E et al. [18] modified the structure and selection mechanism of the random forest algorithm to improve the prediction performance. Authors have verified all the evaluation measure and basing on the feature selection, clustering etc, they have done the voting procedure. The main objective of their work was the combination of the construction and voting method of random forest algorithm. They found the positive effect on the performance by using 24 datasets. Hari Dahal et al. [36] took six soil variables with crop yield data to find the level of crop productivity. They found some of the soil variables have extremely correlated. So to estimate the potency of the relationship they developed the multiple regression models and applied F-Test to know which variable is most significant and found that total nitrogen, organic matter and phosphorous affect the yield of paddy. J. P. Powell et al. [37] analyses the various weather events on the crop winter wheat taking the data on the farm based and of 334 farms for 12 years. They have used the F-Test to find the significance of weather events in the model. They observed and concluded that, the effect of weather events on yield is time specific and also found that the high temperature and precipitation events significantly decrease yields.

Ke Yan et al. [24] studied both the linear and non-linear SVM-RFE algorithm. They have analyzed the correlation bias and anticipated a new algorithm such as, SVM-RFE+CBR. They have implemented in the synthetic dataset. Lastly they found the accuracy on their proposed method. Meng-Dar Shieh et al. [25] proposed one method to eliminate the problem of choosing the features subset. Shruti Mishra et al. [26] recommended one extensive deviation of SVM-RFE and SVM-T-RFE. They found the maximum accuracy in case of classification taking the less subset of gene sets and also of high dimensional data. They have also compared with other two methods such as SVM-T-RFE and SVM-RFE and conclude that the projected step by step method is 40% better than SVM-RFE and 25% better than SVM-T-RFE. The ranking strategies adopted by the above mentioned authors have motivated us to carry forward our research on agricultural and climatic datasets.

## 3 Schematic representation of proposed method

The feature ranking methods are mainly used to rank the features. In this study, a revolutionary effort based on feature ranking methods to find the significant climatic features which affects mostly on the yield of rice of the three coastal districts of Odisha for both the season such as :Rabi and Kharif have been introduced. This empirical study mainly focuses on the selection of significant features through feature ranking and feature fusion based strategies. It works in three important phases, in the first phase known as feature ranking, Random Forest, SVR-FRE and

F-Test based regression methods are explored to rank all the features of the datasets, then in second phase, new ranks have been evaluated by considering all the ranked features from above mentioned ranking techniques and finally, ELM based regressor has been used to empirically evaluate and validate the yield modeling. The Figure 2 illustrates the flow of implementation of proposed ELM based regressor model to obtain the important features that contribute to the yield of rice production in the coastal areas of state of Odisha.

### 3.1 Data set description

The dataset  $D$  is composed of Odisha district of India (Figure 1). Let  $d_i \in D \quad \forall i = 1, \dots, 31$  features that is 31 years of data. where  $|d_i| = 25$  features that is represents the attributes of the datasets. Different parameters are, such as  $p = \{\text{maxtemperature, mintemperature, rainfall, humidity}\}$  that effect the rice production. Since, there are two types of rice production seasons such as; *Rabi* and *Kharif* produced between months 'January–May' and 'June–December', hence  $p_i$  is collected over the range of six months each resulting 24 set of attributes and 25<sup>th</sup> attribute is the production in *hector* of crops for particular year.

The rice production graph for those three coastal areas of Odisha from the year 1983-2014 is shown in **Figure 3(a)** and **Figure 3(b)** for Rabi and Kharif season respectively. The detail description of datasets with standard deviation (Std. Dev.) for three areas is shown in Table 1.

The range and average values of the parameters such as; rainfall in mm/hector, maximum and minimum temperature in °C, mean relative humidity both at 8.30 am and 5.30 pm, of all three datasets with respect to three coastal districts are shown in **Table 2** for Rabi and Kharif seasons.

### 3.2 Study procedures

This section presents a usable scheme to predict the effect of climatic parameters for rice yield in the coastal areas of a state of India, Odisha, during both the Rabi and the Kharif season. These steps are narrated as follows:

- Collection of the raw data including climatologic characteristics and rice production per hector.
- Calculating the range and average of parameters of those datasets for proper knowledge about the features.
- Defining the attributes affecting the rice yield.
- Redefining the datasets and constructing the database of all tuples according to the selected attributes.
- Dividing the raw data into training and testing datasets.
- Designing the feature ranking models to rank all the features of individual datasets for further processing.

- Designing a feature level fusion model using Borda Count to generate a new set of ranked features by taking the ranked features from all three feature ranking strategies for further analysis.
- Designing an ELM based regressor to classify the datasets with the newly ranked features to measure the importance of each feature.
- The accuracy of ELM regressor has been calculated using by R2 score decreasing one by one feature from the datasets.
- Finally, with respect to maximum accuracy, top 5 ranked features are selected, which decide the importance of climatic parameters in rice crop production both Rabi and Kharif in three different districts.
- Finally, with respect to maximum accuracy, top 5 ranked features are selected, which decide the importance of climatic parameters in rice crop production both Rabi and Kharif in three different districts.

## 4 Methodologies adopted for experimentation

This section discusses the various methodologies such as random forest; F-Test and SVR-RFE used for feature reduction and ELM for classification are discussed in this section.

### 4.1 Random forest

Random forest or Random Forest is one of the most important and popular supervised learning algorithm. It can be used both for classification and regression tasks. In this case multiple trees are grown. Then for the classification of a new object based on the attributes, a classification is given by each tree and that is the tree 'votes' for that class. The most votes over all the trees in the forest are chosen for classification and average of outputs by different trees in case of regression. Random forest is one of the ensemble methods of decision trees. Breiman proposed random forest where he adds an extra layer of randomness to bagging [19]. Random forest has a vast number of applications due to its good constancy and simplification [19, 20, 21, 22, 23].

### 4.2 F-Test for regression

The F-Test for linear regression is one of the methods to know the significance of any variable among the independent variables in a multiple linear regression. How the null hypothesis can be tested in a multiple regression model with intercept can be described by the F-Test for regression [27, 28].

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0 \quad (1)$$

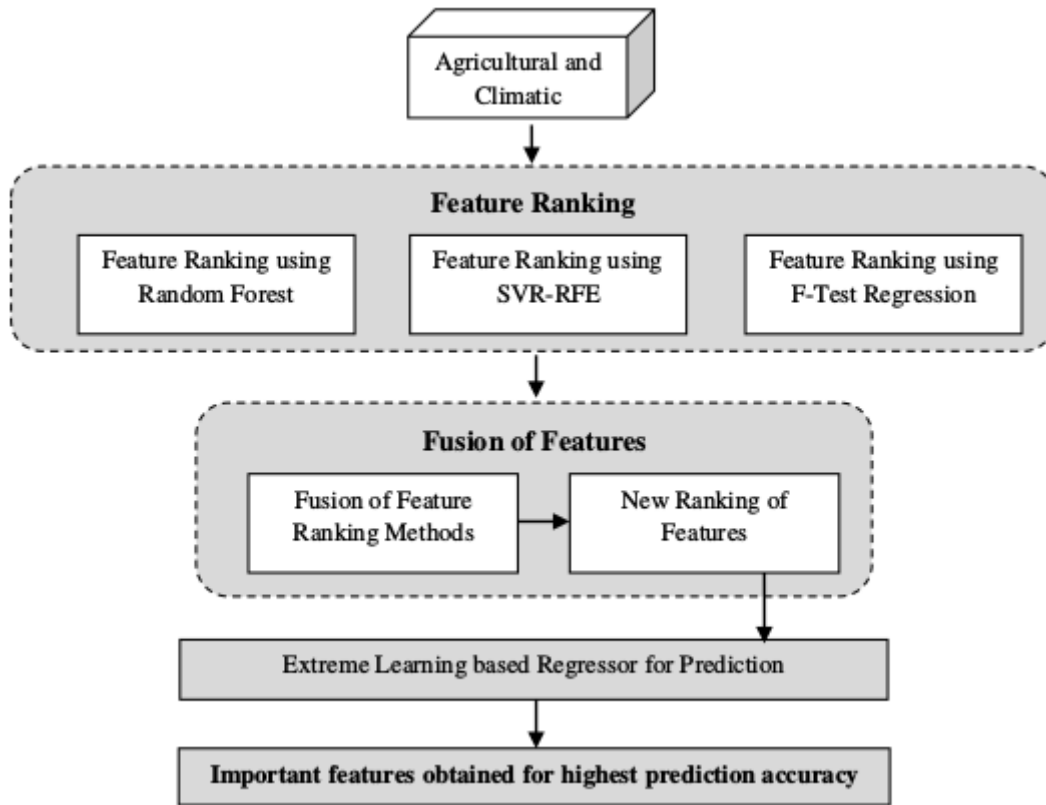


Figure 2: Graphical abstract of proposed model.

Table 1: Description of real datasets collected over period 1983-2014 for Rabi and Kharif production.

Seasons	Rabi			Kharif		
	Dimension	Mean	Std. Dev.	Dimension	Mean	Std. Dev.
Balasore	31 × 25	47.8386	20.84	31 × 35	81.6430	43.7791
Cuttack	31 × 25	44.7391	18.43	31 × 35	80.6577	50.6339
Puri	31 × 25	47.6373	25.77	31 × 35	78.9684	44.2095

$$H_0 : \beta_i \neq 0 \text{ for atleast one value of } i \quad (2)$$

Then, assuming the null hypothesis as true we have to test.

$$F = \frac{MSM}{MSE} = \frac{\text{Explained Variance}}{\text{Unexplained Variance}} \quad (3)$$

Where,  $MSM = \frac{SSM}{DFM}$  and  $MSE = \frac{SSE}{DFE}$   
 MSM=Mean Squares for Model  
 SSM=Corrected Sum of Squares of Models  
 DFM=Corrected Degrees of Freedom for Models  
 DFE=Degree of Freedom for Error

Then, using an F-table or statistical software, we have to find confidence interval for degrees of freedom.

### 4.3 Support vector regressor-recursive feature elimination (SVR-RFE)

SVR-RFE is one of the variable selection or feature selection method. It is an optimization method for finding the best performing feature set. Repeatedly it creates models taking features subset and next with left features and lastly it ranks the features on the basis of order of elimination [24-26]. First the algorithm is trained by SVM with a linear kernel and then the features are detached recursively using the smallest ranking criterion. In order to generate a rank the weight vector needs to be calculated as given in Equation (4).

$$W = \sum_{i=1}^n \beta_i x_i y_i \quad (4)$$

Where,  $i$  is the number of features ranging from 1 to  $n$ ;  $\beta_i$  is the Lagrangian Multiplier estimated from the training set;

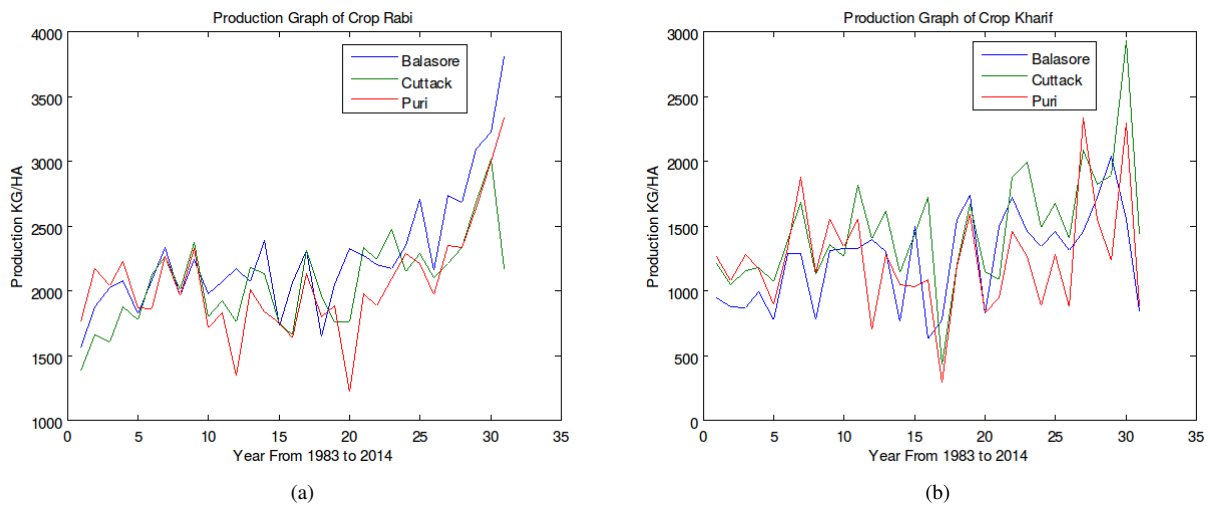


Figure 3: Graphical representation of rice production of three regions for Rabi and Kharif seasons.

Table 2: Range and average values of the parameters in datasets.

Districts	Parameter	Range		Average		Average Rice Production	
		Rabi	Kharif	Rabi	Kharif	Rabi	Kharif
Balasore	Rainfall(mm/hector)	0.0–431.2	0.0–696.5	55.2	280.3		
	Max Temperature (°C)	25–42	13–37.4	33	32		
	Min Temperature (°C)	9.8–32	11.9–28	21	25	2261.50	1243.8
	Mean Relative Humidity at 8.30AM (%)	53–81	35–88	68	79		
	Mean Relative Humidity at 5.30PM (%)	45–87	34–89	66	78		
Cuttack	Rainfall(mm/hector)	0.0–477.8	0–752.8	36.42	268.2		
	Max Temperature (°C)	26–40	26.8–38	31.76	32		
	Min Temperature (°C)	11–32	15–33	20.92	25	2064.71	1472.5
	Mean Relative Humidity at 8.30AM (%)	58–95.5	67–95.4	84.33	87		
	Mean Relative Humidity at 5.30PM (%)	29.3–89	12–90	50.27	73		
Puri	Rainfall (mm/hector)	0.0–735.5	0.0–826.5	27.13	247		
	Max Temperature (°C)	25–35.3	20.8–40.8	30.43	32		
	Min Temperature (°C)	12–29	15.2–29	23.49	26	2053	1240
	Mean Relative Humidity at 8.30AM (%)	70–92	66–92	80.74	83		
	Mean Relative Humidity at 5.30PM (%)	64–90	17–91	78.87	81		

$x_i$  is the gene expression vector for sample  $i$  and  $y_i$  is the class label of  $i$  ( $y_i \in [-1, +1]$ )

#### 4.4 Extreme Learning Machine (ELM)

Artificial Neural Network (ANN) is one of the best examples of classification and regression technique which works on back-propagation method. In this case weights are adjusted by trial and error methods. But there are various disadvantages of ANN, such as; *local minima*, *over fitting*

*problem* and *large training time* [38–40]. To overcome the problem of memory requirements, Hung et al. [29] projected new method which is based on the least square algorithm for classification and regression problem, known as ELM. ELM also has unique minimum solution, with both *smallest training error* and *smallest weight norm*, *does not need a stopping methods*.

ELM is a learning neural algorithm, introduced to develop the efficiency of Single Layer Feed Forward Neural Network (SLFN). This section will briefly explain the

---

**Algorithm 1:** SVR-RFE [[21, 22, 23]]

---

**Input:** Initial feature subset,  $F = \{1, 2, \dots, n\}$

**Output:** Rank list according to smallest weight criterion,  $R$ .

- 1 Set  $R = \{\}$
  - 2 Repeat 3 -8 until  $F$  is not empty
  - 3 Train the SVM using  $F$ .
  - 4 Compute the Weight Vector using (1)
  - 5 Compute the Ranking Criteria,  $Rank = W^2$
  - 6 Rank the features as in sorted manner,  
 $New_{Rank} = Sort(Rank)$
  - 7 Update the Feature Rank list  
 $Update \quad R = R + F(New_{Rank})$
  - 8 Eliminate the feature with smallest rank  
 $Update \quad F = F - F(New_{Rank})$
- 

working principle of ELM [30, 31, 32].  $N$  is given as a training sample, where  $(X_i, Y_j) \in R^n \times R^m$ . Here,  $j = 1, 2, \dots, N$  and the number of hidden nodes is considered as  $M$ . Representing the output of SLFN, the equation is formulated in (5).

$$output_k = \sum_{j=1}^M \beta_j f(X_k) = \sum_{j=1}^M \beta_j f(X_k; a_j, b_j), \quad k = 1, 2, \dots, N \quad (5)$$

Where, with respect to the input sample, the output vector is  $output_k$  and  $f(X_k; a_j, b_j)$  is the activation function.  $a_j$  and  $b_j$  are the randomly generated learning parameter of the  $k^{th}$  hidden node and (5) can be compactly written as

$$H \times \beta = Calculated \ Output \quad (6)$$

Here,

$$H = \begin{bmatrix} f(a_1.x_1 + b_1) & \dots & f(a_M.x_1 + b_M) \\ \vdots & \ddots & \vdots \\ f(a_1.x_N + b_1) & \dots & f(a_M.x_N + b_M) \end{bmatrix}_{N \times M}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times 1}$$

$$Calculated \ Output = \begin{bmatrix} Output_1^T \\ \vdots \\ Output_N^T \end{bmatrix}_{N \times 1}$$

Where,  $H$  is the output matrix, (2) can be linear system by analytically determine the output weights by finding the least square solution, which is defined in (3)

$$\hat{\beta} = inv(H' \times H) \times H' \times trainoutput \quad (7)$$

Where,  $trainoutput$  is the output of the training data and the benefit of the ELM is that, the output weight is systematically calculated by using some mathematical transformation, avoiding the lengthy process of training and simultaneously no iterative adjustment of the training parameter is required.

### 4.5 Fusion strategies

The Borda Count [41, 42] is one of the superior voting system. In this case the voters rank the candidates according to the inclination. Then the points are formed from ranking. The candidates which will gate score one point then ranked last, then score two and next-to-last and so on. Who will secure the more points then declared as winner. There are various other standard voting systems such as: Alternative vote and the single transferable vote, but the advantages of Borda count are, all the MPs have the support of a majority of their votes. The parties nominate the good one. This method is a kind of group consensus functions which maps the inputs of individual rankings to a combined form of ranking which leads to a most appropriate and relevant decision making process. With respect to machine learning, Borda Count is defined as a sum of number of classes ranked below the class by each classifier. The degree of the Borda Count reflects the level of agreement that the input pattern belongs to the considered class. The main advantage of this method is to implement and does not require any training.

### 4.6 Validation strategies adopted

$R^2$  is one of the statistical compute to find the fitness of the regression line with the data [43]. Some knowledge regarding the goodness of fit of a model can be defined by this statistic [35, 36]. A linear model explains the proportion of response variable variation and values of  $R^2$  always lie between 0 and 100% or 0 and 1, where; 0% or 0 indicates that the model explains none of the variability of the response data around its mean and 100% or 1 indicates that the model explains all the variability of the response data around its mean and this statistics measure of how well the regression predictions approximate the real data points. An  $R^2$  of 100% or 1 indicates that the regression predictions perfectly fit the data.

## 5 Experimentation and model evaluation

### 5.1 Experimental setup

In this work all the implementations have been carried out using python programming environment in Linux operating system with a minimum hardware configuration of 4GB RAM and 100GB hard disk. First of all, the different activation functions are tested for best suitability to our prob-

lem domain. Then, different feature ranking strategies have been tested with ELM. Finally, the proposed fusion of feature ranking has been tested. The parameters used for experimentation is illustrated in Table 3.

## 5.2 Parameters used

The **Table 3** gives the details of the parameters used for the implementation.

## 5.3 Feature ranking methods

Here three different feature ranking methods such as Random Forest, SVR-RFE and F-Test have been experimented for regression. In literature, it has been found that, these are mainly used for ranking of genes in gene expression datasets and in this study; the same methods are used to rank the features of rice crop prediction datasets. This methodology works in three different steps such as; (a) first, the three ranking algorithms outputs three *different ranks* to each feature of the dataset; (b) secondly, a *feature fusion* method based on *Borda Count* has been used to evaluate the final rank of each feature and; (c) finally, these *newly ranked features* are evaluated by ELM based regressor to measure the *importance of each feature*. The accuracy of ELM regressor has been calculated by decreasing one by one feature from the datasets. Finally, with respect to maximum accuracy, top five ranked features are selected, which decide the importance of climatic parameters in rice crop production both for the Rabi season and the Kharif season in all the districts taken for the analysis. **Figure 4** and **Figure 5** shows the features are arranged in the descending of their  $R^2$  scores measuring the importance of the features after applying the Random Forest feature ranking method on both Rabi and Kharif seasons respectively for Balasore, Cuttack and Puri districts. From **Figure 4** for Rabi season it can be observed that, the features 21, 18, 13 and 11 are having approximate importance scores from 0 to 13, whereas features 7 and 12 are having very less importance scores and rest are in a moderate stage for Balasore district, for Cuttack district, features 0 (first feature) and 7 are having approximate importance scores from 0 to 14, whereas, features 3, 17 and 9 are having very less importance score. Similarly, for Puri district feature 21 has very high importance and 19, 17, 23, 20, 8, 15, 22, 18 and 7 are having moderate scores. Rest others can be ignored due to their very less scores of importance.

Similarly, for Kharif season, from **Figure 5** it can be seen that, the feature 5 is showing highest importance score of 8 and the feature 5 is having the lowest score of importance and rest are lying within the range of 2-6 scores for Balasore district. For Cuttack district, features 1, 24, 8, 9, 23, 14 and 7 are having approximate importance scores from 0 to 7, rest other features are having very less importance scores. Similarly, for Puri district features 8 and are having very high importance with the scores 0 to 16, and 4, 10 and 9 are having moderate scores. Rest others can be ig-

nored due to their very less scores of importance. **Figure 6** and **Figure 7** shows the features with respect to their  $R^2$  scores measuring the importance of the features after applying the SVR-RFE feature ranking method on both Rabi and Kharif seasons respectively for Balasore, Cuttack and Puri districts. From **Figure 6** for Rabi season it can be observed that, the feature 23 is having the 1st rank, then features 15, 9, 21 and 14 are showing better rank and few more are showing moderate rank and feature 4 is having the lowest rank giving rise to non-significant feature. The feature 7 is having the highest rank, and feature 17 is with lowest rank in Cuttack district. Similarly, the feature 19 has very high rank and features 17, 11, 15, 23 are having better rank and feature 4 has less importance in Puri district. Similarly, in **Figure 7**, the feature 27 is experiencing the highest rank, feature 25 and 9 is next to best and feature 0 (first feature) is having less rank with less impact of the feature in Balasore district. For Cuttack district feature 16 is of great importance and feature 34 is of no or less importance, therefore can be ignored. Feature 29 is showing the highest rank and 23, 9, 8 and 20 features are also experiencing better scores, but feature 33 is with the lowest rank in Puri district. The importance of features for both Rabi and Kharif seasons using F-Test for regression has been plotted in **Figure 8** and **Figure 9** respectively. From the experimentation of Rabi season (**Figure 8**), it can be seen that, for Balasore district feature 21 is with the highest score, features 22,24,8,7,5,0 are with lowest scores,4,13,19 are negligible score and rest others are having moderate scores. For Cuttack district feature 6 is with the highest score, 1, 5, 8, 9, 13, 17 and 18 are of no importance and they do not contribute for processing. Similarly for Puri district feature 17 and 21 are having the highest scores, features 1,4,5,13 and 23 are with lowest scores and also it can be seen that rest other features are also not showing better scores. From **Figure 8** for Kharif season, the features 16, 1 and 8 are having the highest importance for Balasore, Cuttack and Puri districts respectively. Features 7, 8, 12, and 31 for Balasore, 2, 5, 6, 10 and 12 for Cuttack and 6, 15, 20, 24, 28 and 31 for Puri datasets are showing scores of least importance.

## 5.4 Fusion of feature ranking methods

Here, a multiple ranking fusion scheme has been proposed. In this scheme, the individual rankings using different ranking methods have been obtained and then those ranked features are combined to obtain the final rankings of features. The most popular and effective method for fusion used here is Borda count method.

Mathematically, the fusion of features based strategy can be proposed as; let the dataset is defined as  $DS = \{x_1, x_2, x_3, \dots, x_n\}$ , where  $x_1, x_2, x_3, \dots, x_n$  represents  $n$  number of features of the dataset and  $r_1, r_2$  and  $r_3$  are three ranking methods used and the proposed fusion of ranking strategy can be described as shown in **Figure 10**. The importance of features for both Rabi and Kharif seasons using fusion of ranking strategy for regression has



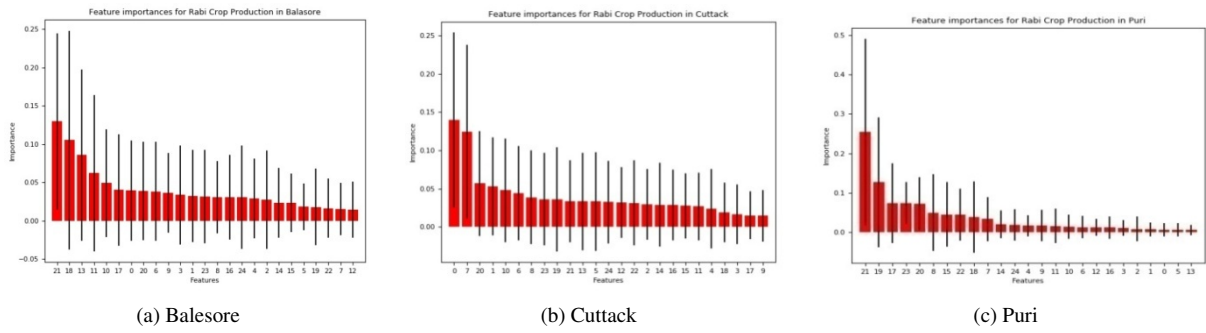


Figure 4: Feature ranking using Random Forest for Rabi season in three districts.

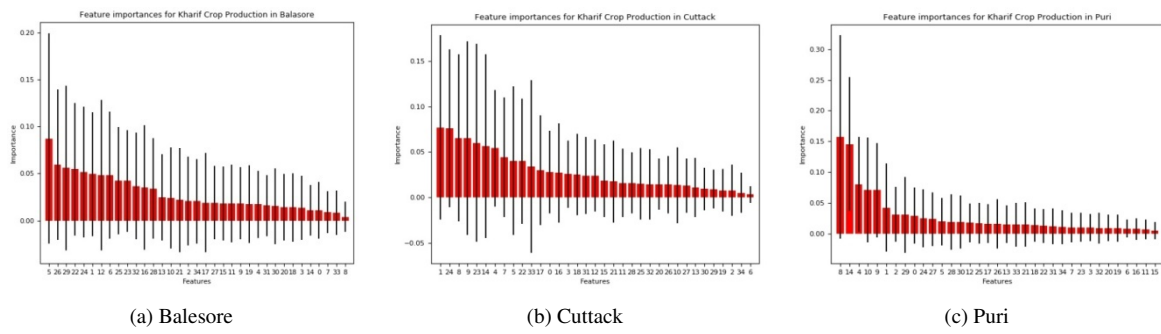


Figure 5: Feature ranking using Random Forest for Kharif season in three districts.

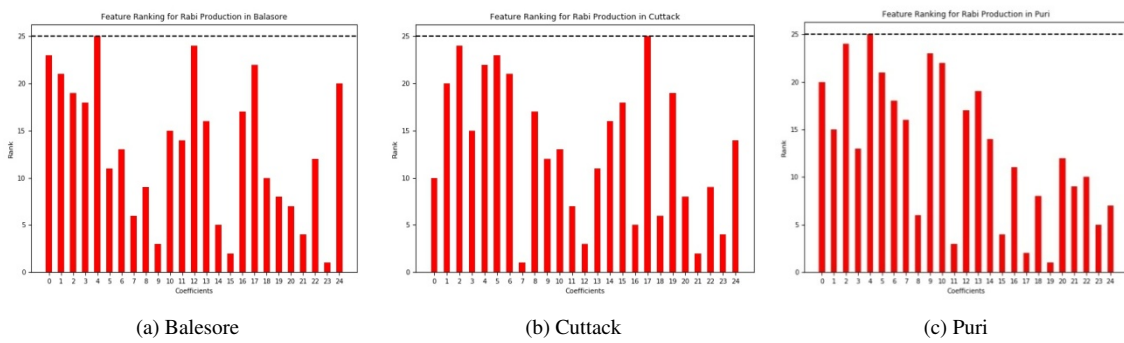


Figure 6: Feature ranking using SVR-RFE for Rabi season in three districts.

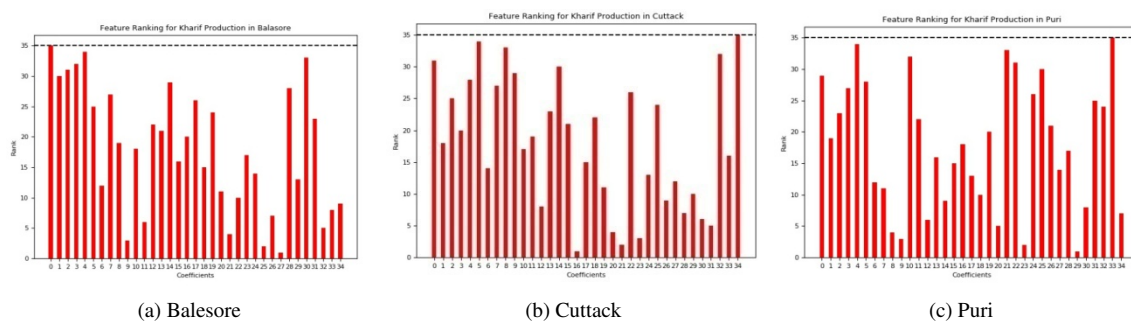


Figure 7: Feature ranking using SVR-RFE for Kharif season in three districts.

Table 3: Parameter set up for ranking methods.

Techniques	Parameters
Random Forest for feature ranking	No of estimators=1000, criterion=mean square error
SVR-RFE for feature ranking	C=1.0 (Penalty parameter), Base estimator=SVR, kernel=linear, no of features to select=1, step=1
F-Test for feature ranking	Score_function=Ftest, no of features=1
Extreme Learning Machine	No. of hidden layers - 500, Activation function - Multi-quadratic

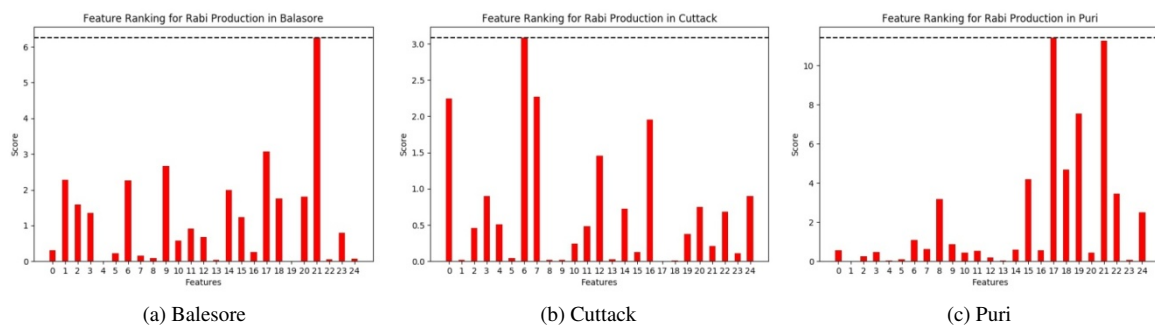


Figure 8: Feature ranking using F-Test for Regression for Rabi season in three districts.

been plotted in **Figure 11** and **Figure 12** respectively and the five top ranked features obtained are listed in **Table 4**.

### 5.5 Extreme learning machine regressor

In this work, first, all the variants of ELM regressors have been evaluated with different activation functions such as; tanh, sine, tribas, inv-tribas, sigmoid, hardlim, softlim, gaussian, multiquadratic, inv-multiquadratic etc. Among these functions it has been observed that, tribas, inv-tribas, hardlim, softlim and Gaussian functions gives a negative value of R2 score and score of tanh, sine, sigmoid, multiquadratic and inv-multiquadri functions are found to be  $\geq 98\%$  as detailed in Figure 13 and Figure 14 and also Table 5 and Table 6, shows the graph for R2 score for different activation functions for ELM to predict Rabi and Kharif rice crops respectively. From all those ten activation functions multiquadratic is having the highest R2 score while considering all the districts for Rabi and Kharif seasons. Hence, for the experimentation, mutiquadric function has been considered.

### 5.6 ELM-Regressor for varying number of features

Once, the newly ranked features are obtained from proposed feature fusion strategy and the activation function (multiquadratic) have been also found to be used by ELM, now the accuracy of ELM Regressor has been calculated by decreasing one by one feature from the datasets as shown in Figure 15 and Figure 16.

Table 7 and Table 8 depicts the accuracy of prediction obtained by multiquadratic based ELM regressor for Rabi and Kharif seasons respectively for all three coastal regions by decreasing the features one by one. The maximum number features those shows above 99% accuracy are coded in red, green and blue colors for Balasore, Cuttack and Puri districts respectively for proper visualization of the readers. From Table 7, it is evident that, while decreasing the number of features from 25 to 20, 15, 14, 13, 12, 8, 6 and even 3 shows above 99% prediction accuracy for Balasore, for Cuttack the number features showing 99% prediction accuracy are 20, 10, 9 and, similarly, for Puri, 18, 15, 11, 10, 6, 3 and 2 number of features are giving maximum prediction accuracy above 99%.

Similarly, from Table 8, it can be observed that, while decreasing the number of features from 35 to 34, 33, 30, 26, 22, 23, 20, 18, 17, 16 and 15 shows above 99% prediction accuracy for Balasore, for Cuttack only 18, 11, 6, 4, 3, 2, and 1 number features are below 99% prediction accuracy and rest are giving above 99%, and, similarly, for Puri, 33, 30, 27, 26, 24, 23, 18, 16, 11, 10, 9, 8, 7, 4, 3, 2 and 1 number of features are giving below 99% prediction accuracy. From those two table and figures this, it can be accomplished that, to predict the crop yield for Rabi season less number of features are working better in comparison to Kharif seasons.

### 5.7 Result analysis

After obtaining the top five ranked features and the varying number features which give above 99% prediction ac-

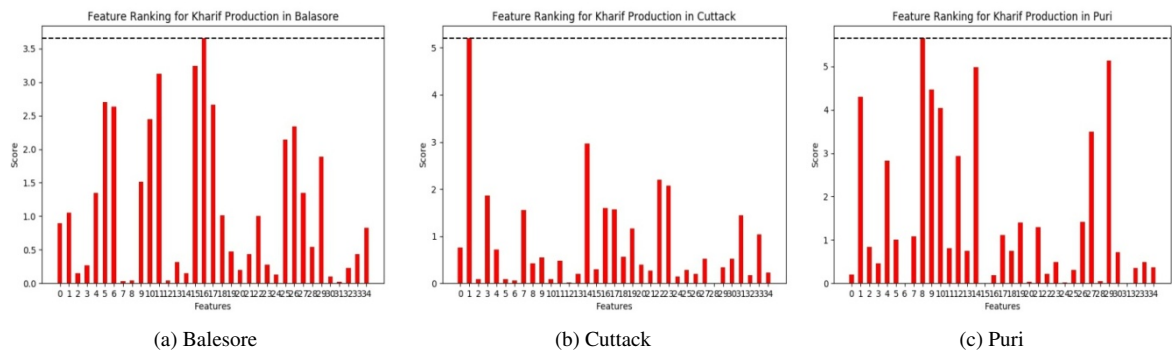


Figure 9: Feature ranking using F-Test for Regression for Kharif season in three districts.

Table 4: Five top ranked features extracted using feature ranking based on Borda Count feature fusion strategy of three districts of Rabi and Kharif season.

Seasons	Balesore		Cuttack		Puri	
	Feature No	Feature Name	Feature No	Feature Name	Feature No	Feature Name
Rabi	23	May RH-8:30 AM	12	Mar Min Temp	22	Apr RH-5:30 PM
	15	Jan RH-8:30 AM	21	Apr RH-8:30 AM	24	May RH-5:30 PM
	9	May Max Temp	7	Mar Max Temp	11	Feb Min Temp
	11	Feb Min Temp	23	May RH- 8:30 AM	15	Jan RH 8:30 AM
	14	May Min Temp	16	Jan RH-5:30 PM	23	May RH-8:30 AM
Kharif	27	Sep RH-8:30 AM	31	Nov RH-8:30 AM	12	Nov Max Temp
	25	Aug RH-8:30 AM	21	June RH-8:30 AM	19	Nov Min Temp
	9	Aug Max Temp	23	July RH-8:30 AM	9	Aug Max Temp
	21	June RH-8:30 AM	20	Dec Min Temp	25	Aug RH-8:30 AM
	26	Aug RH-5.30 AM	16	Aug Min Temp	20	Dec Min Temp

Table 5: R<sup>2</sup> score of all activation functions of ELM for Rabi seasons.

ELM Activation Functions	R <sup>2</sup> score for Rabi Season		
	Balesore	Cuttack	Puri
TANH	0.998093743193	0.994257107884	0.9989356101
SINE	0.996717695318	0.99942453749	0.983896079504
SIGMOID	0.987233114958	0.998330563403	0.999426924698
<b>MULTIQUADRIC</b>	<b>0.999957522834</b>	<b>0.999818219303</b>	<b>0.999726152755</b>
INV-MULTIQUADRIC	0.958613787069	0.935259681129	0.966708028068
TRIBAS	-12.4064777143	-11.4144046567	-9.03257377773
INV-TRIBAS	0.0	-2.22044604925e-16	-2.22044604925e-16
HARDLIM	0.0	-2.22044604925e-16	-2.22044604925e-16
SOFTLIM	0.0	-2.22044604925e-16	-2.22044604925e-16
GAUSSIAN	-1.13177301381	-0.59754939441	-0.0727264050254

curacy for both the seasons, in this section an attempt has been made to validate proposed fusion of feature ranking based strategy with Random Forest, SVR-RFE and F-Test

with multiquadratic based ELM to find the impact of fusion based strategy with non-fusion based ranking strategies for the maximum number features that contribute to achieve

Table 6: R<sup>2</sup> score of all activation functions of ELM for Kharif seasons.

ELM Activation Functions	R <sup>2</sup> score for Kharif Season		
	Balesore	Cuttack	Puri
TANH	0.999802124998	0.900092462207	0.941367554859
SINE	0.981838265602	0.983905092261	0.854629459493
SIGMOID	0.993512967504	0.936873558516	0.964438947941
<b>MULTIQUADRIC</b>	<b>0.999993905565</b>	<b>0.999624110222</b>	<b>0.991070886794</b>
INV-MULTIQUADRIC	0.979667648088	0.999512861069	0.960984673885
TRIBAS	-21.7579913615	-37.3921891299	-8.74341923183
INV-TRIBAS	0.103557054691	-4.4408920985e-16	0.0669379515163
HARDLIM	0.0	-4.4408920985e-16	-8.881784197e-16
SOFTLIM	0.103557054691	-4.4408920985e-16	0.0669379515163
GAUSSIAN	-0.308828188795	-1.77193827293	0.549504155107

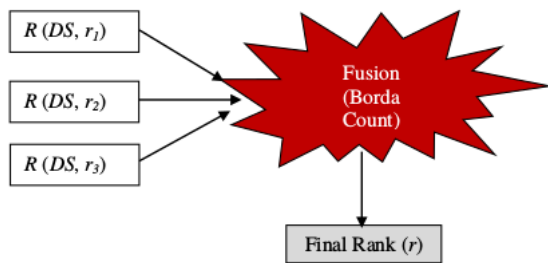


Figure 10: Fusion of feature ranking strategy.

99% prediction accuracy as shown in Table 9 and Table 10 for Rabi and Kharif season crops.

For Rabi season crop from Table 9, it can be seen that, proposed fusion based ranking strategy when compared with non fusion based strategies, the maximum number of features that contribute predictive accuracy above 99% for ELM with Random Forest is 7, 10, 6; ELM with SVM-RFE is 5, 9, 4 and similarly ELM with F-Test needs 9, 11, 8 numbers of features to give 99% and above predictive accuracy. While with a very less number of features such as; 3, 5 and 2 can predict above 99% accuracy for Balesore, Cuttack and Puri districts respectively. From Table 4, where the top five ranked features extracted from fusion strategy, it can be concluded that the crop yield for Balesore district in Rabi season can be accurately predicted if we consider only three features out of RH at 5.30 PM of March, April, May, RH of February 8.30 AM and 5.30 PM, because they are affecting the rice crop yield maximum. The five features that affect the rice yield during Rabi season for Cuttack district are; RH of March, April and May and also the minimum and maximum temperature of May month. Similarly, the two features that affect the crop yield of Puri district during Rabi season are out of five features such as; RH of March and May months and minimum temperature of March and May months. From this observation, it can be said that the features containing RH in 8.30 AM

Table 7: Performance of ELM with varying number of features for Rabi crop prediction.

No. of Features	Balesore	Cuttack	Puri
25	0.9721452943	0.983009018	0.8918569778
24	0.8820921735	0.9211651663	0.899522749
23	0.9723284733	0.9717225517	0.8644802695
22	0.9844701984	0.9800205232	0.9897965576
21	0.9668404406	0.9551503977	0.9665234947
20	0.9996177026	0.9999348622	0.9869710443
19	0.9592805399	0.9127841794	0.9398241394
18	0.8356003816	0.9081500374	0.9942342785
17	0.9511241577	0.9780000307	0.9288099884
16	0.9354752886	0.9358388115	0.9363188192
15	0.9930751632	0.9172893274	0.9928662122
14	0.9901838183	0.9617978239	0.9512236619
13	0.9999946834	0.9896162607	0.9585303661
12	0.9934721511	0.9027066465	0.9372090401
11	0.9594424161	0.9510466357	0.9919344792
10	0.8894488099	0.9943953688	0.992055051
9	0.9765177632	0.999323231	0.971105448
8	0.9990643405	0.9784069021	0.9623709905
7	0.9735100076	0.9850878134	0.9978247397
6	0.9968633499	0.9728457206	0.9757838604
5	0.9135013909	0.9969514165	0.9785948706
4	0.9815795296	0.836152001	0.9720037388
3	0.9992149872	0.9126616196	0.998992344
2	0.9183391785	0.9892897973	0.9945608993
1	0.3091946087	0.7773622128	0.5590549638

and 5.30PM are the mostly affecting rice crop yield in all the three districts for the Rabi season crop.

Similarly, while analyzing the Table 10 for Kharif season for all the district datasets, the observation says, Kharif season crops needs more parameters or features to be considered in comparison to Rabi season crops which is evident from Table 8 and Table 10. The top 15, 5 and 5 ranked features are need to accurately predict the rice yield during this season for Balesore, Cuttack and Puri districts respectively. Observing from Table 4, it can be established that, for Balesore district 15 numbers of features are affecting

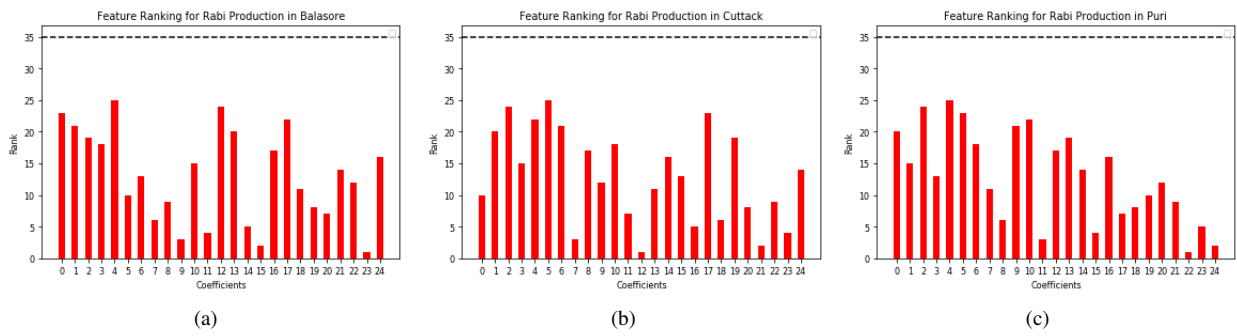


Figure 11: Feature ranking based on Borda Count based feature fusion strategy for Rabi season in three districts.

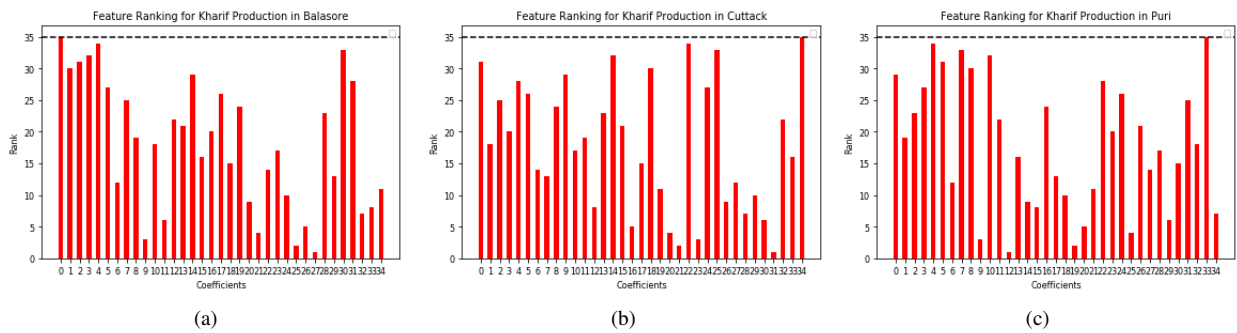


Figure 12: Feature ranking based on Borda Count based feature fusion strategy for Kharif season in three districts.

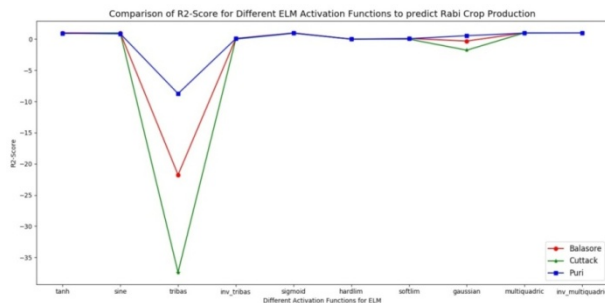


Figure 13: Performance comparison of different activation functions for ELM for Rabi Crop prediction in three different districts.

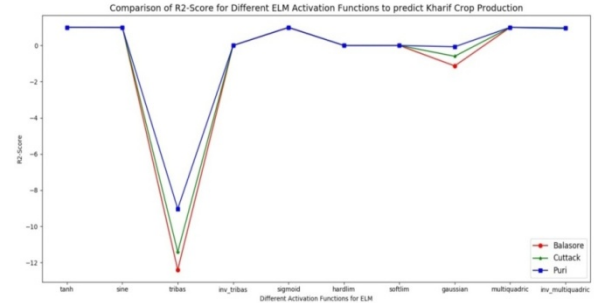


Figure 14: Performance comparison of different activation functions for ELM for Kharif crop prediction in three different districts.

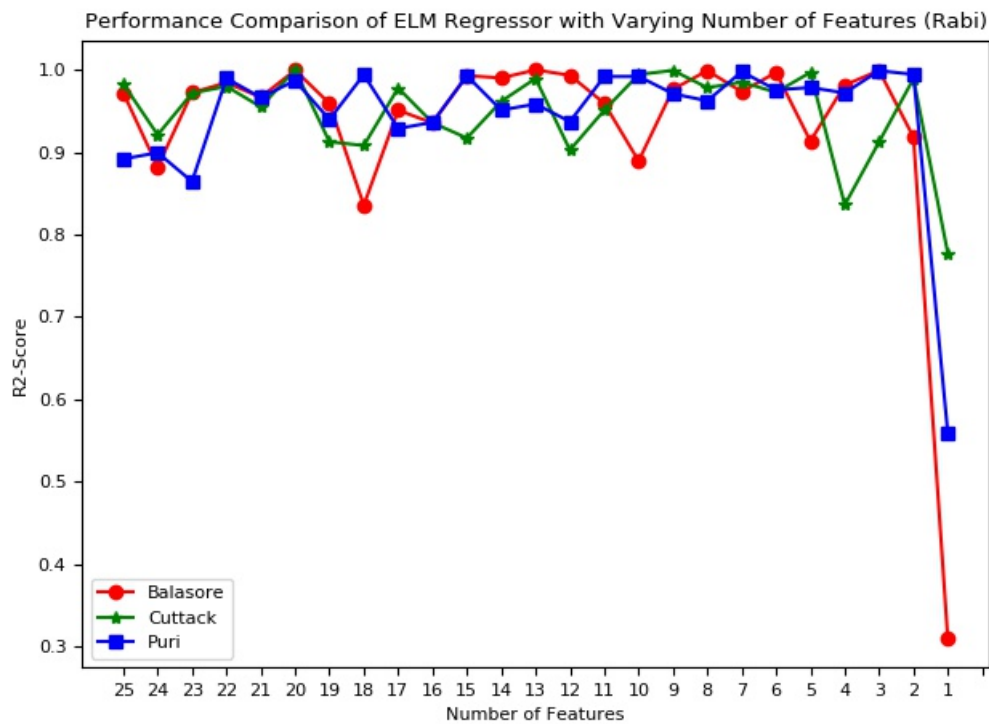


Figure 15: Performance comparison of ELM based Regressor for rice crop prediction (Rabi season) with varying number of features.

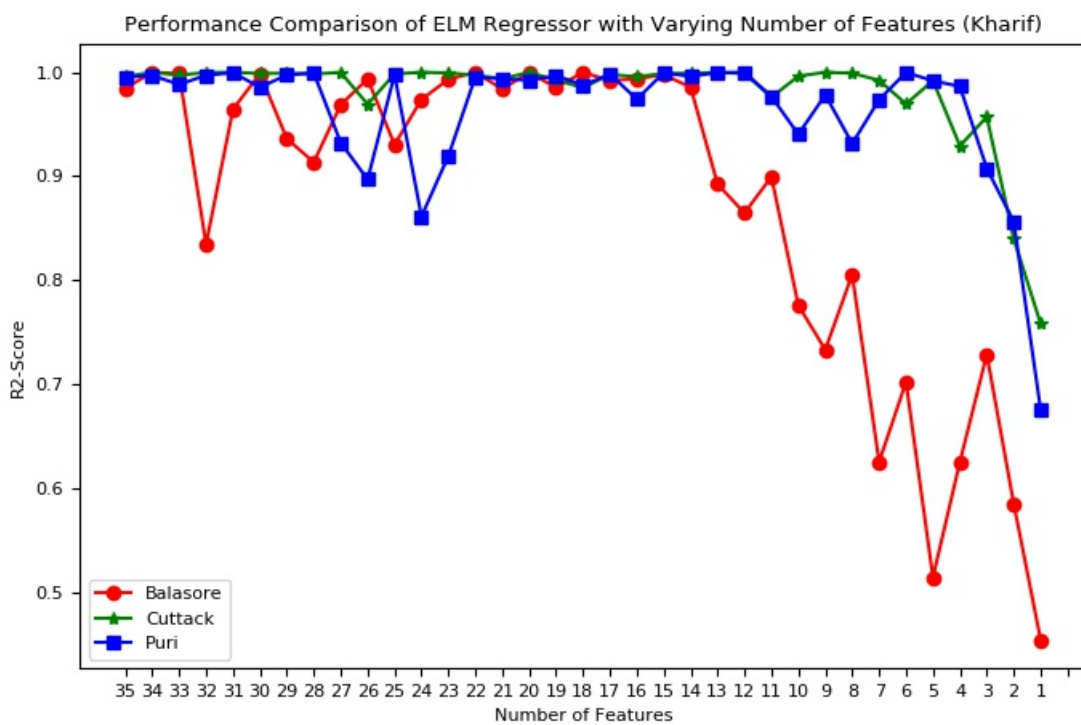


Figure 16: Performance comparison of ELM based Regressor for rice crop prediction (Kharif season) with varying number of features.

Table 8: Performance of ELM with varying number of features for Kharif crop prediction.

No. of Features	Balesore	Cuttack	Puri
35	0.9846787896	0.9957370161	0.9951327585
34	0.9998355712	0.9999942814	0.9968069715
33	0.9988511032	0.9971118552	0.9882099882
32	0.834652697	0.9997670449	0.9970224703
31	0.9644203951	0.9999974004	0.9995567779
30	0.9985549641	0.998758502	0.9855062635
29	0.9361953426	0.9990206928	0.9973169474
28	0.9134833525	0.9982784062	0.9993592848
27	0.9687297599	0.9997739351	0.9323478922
26	0.9935929806	0.9691971505	0.8970090564
25	0.9303826554	0.9983009945	0.9976403045
24	0.9738004793	0.9999393614	0.8613974874
23	0.9931366576	0.9993454567	0.9194678778
22	0.9998339948	0.9970860103	0.9953840259
21	0.9838072021	0.9938437106	0.9934978926
20	0.9998996937	0.9997478388	0.9923063525
19	0.985075577	0.9925039823	0.9966797975
18	0.9999812113	0.9850548875	0.9875049342
17	0.991885116	0.9977728244	0.9982482335
16	0.9937114339	0.9956994688	0.9718,40716746
15	0.9975411687	0.9987271376	0.9994908157
14	0.9855317118	0.9991787244	0.9958694953
13	0.8924205418	0.9997424156	0.9996766792
12	0.8646878928	0.9986409272	0.9999242476
11	0.8996188387	0.9760167761	0.9766086922
10	0.7759922185	0.9961673342	0.9402894268
9	0.7333665426	0.9999887705	0.9782569996
8	0.8055817301	0.9990313839	0.9314026123
7	0.6255842427	0.9922110063	0.9736563656
6	0.7024204135	0.9695098805	0.9997970624
5	0.5146431718	0.9919755462	0.9917328613
4	0.6254793909	0.9292500719	0.9863601838
3	0.7288521573	0.9570653424	0.9075062814
2	0.5854590008	0.83996043	0.8564995075
1	0.4545559232	0.7583238157	0.67564097

the crop yield out of while top five features such as; RH of October, November, December during 8.30 AM and 5.30 PM are shown due to less space. The features affecting the Cuttack district rice yield are RH of July, Sept and October during 8.30 AM and 5.30 PM and also the minimum temperature during September and November months; for the Puri district, the 5 features that affects the rice yield are RH of June, August, September and December mostly 5.30 PM and only 8.30AM in December and also the minimum temperature during October months. From this, it can be concluded that, the features affecting mostly for rice yield are RH during 8.30 AM and 5.30 PM during Kharif season for all three districts as similar to Rabi season.

### 5.8 Statistical validation

Paired T-test is one of the methods, to assess the consequence of the proposed fusion of feature ranking approach. The outcome produced by ELM-SVR-RFE was compared with proposed approach for five independent runs considering top five ranked features. Here, only ELM-SVR-RFE for statistical validation has been considered for paired test,

as it gives better result than the other basic feature ranking based methods. There is no difference found between the outcomes of the two methods that the null hypothesis was the case. The outcomes shown both for the Rabi and the Kharif seasons respectively in the Table 11 and Table 12. From the below tables we can see that, the null hypothesis is rejected and average p-value is 0.0023, 0.0021, 0.0044 for the taken three districts such as: Balasore, Puri and Cuttack of Rabi season and 0.0335, 0.0221 and 0.0450 for Kharif season of all three districts such as: Balasore, Puri and Cuttack. We can observe that the values are closer to zero and for this reason the arguments are strengthened and the projected fusion of feature ranking approach has improved performance than the other only feature ranking based methods.

## 6 Discussion on principal findings

The principal aim of the present study is to discover the features those have important role or affects mostly in rice crop production both for the Rabi and Kharif seasons of Balasore, Cuttack and Puri. To obtain our desired result, a fusion based strategy based of feature ranking methods has been proposed and explored. This methodology works in three computational phases and not only finds the most significant features contributing towards rice yield but also shows 99% and above prediction accuracy. According to the results obtained the following are few observations made on this study:

- First, the raw data including climatologic characteristics and rice production per hector are collected for three districts and two seasons and the range and average of parameters of those datasets are computed to have a greater insight about the features for proper understanding.
- The importance of features have been evaluated and those features are selected for prediction of rice yield using, ranking of features by applying Random Forest, SVR-RFE and F-Test ranking strategies. These feature ranking models, rank all the features of individual datasets for further processing.
- A feature level fusion model using Borda Count has been explored to generate a new set of ranked features by taking the ranked features from all three feature ranking strategies for further analysis. From this, top five ranked features contributing mostly for rice yield have been listed in **Table 4**.
- Multiquadratic activation has been confirmed from ten activations functions based on R2 score to be used by the ELM regressor to obtain the rice yield prediction above 99% predictive accuracy by decreasing the features one by one for two seasons and three district datasets and results are shown in **Table 7** and **Table 8**.

Table 9: Performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Rabi crop prediction.

Districts	Number of top ranked features required to achieve a threshold accuracy of 99%			
	ELM with Random Forest	ELM with SVR-RFE	ELM with F-Test	ELM with Proposed Fusion Strategy
Balasore	7	5	9	3
Cuttack	10	9	11	5
Puri	6	4	8	2

Table 10: Performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Kharif crop prediction.

Districts	Number of top ranked features required to achieve a threshold accuracy of 99%			
	ELM with Random Forest	ELM with SVR-RFE	ELM with F-Test	ELM with Proposed Fusion Strategy
Balasore	21	15	24	15
Cuttack	17	10	21	5
Puri	17	12	22	5

Table 11: Paired T-test of Rabi season datasets (all three districts) for the ELM-SVR-RFE approach and proposed Fusion based feature ranking strategy.

Runs	Balasore District Dataset		Puri District Dataset		Cuttack District Dataset	
	Hypothesis Test	p-Value	Hypothesis Test	p-Value	Hypothesis Test	p-Value
	1	1	0.002374351	1	0.002145848	1
2	1	0.002376581	1	0.002763544	1	0.00423645
3	1	0.002432856	1	0.002658974	1	0.00445726
4	1	0.002743567	1	0.002738465	1	0.00465187
5	1	0.002267655	1	0.002748983	1	0.00435478

Table 12: Paired T-test of Kharif season datasets (all three districts) for the ELM-SVR-RFE approach and proposed Fusion based feature ranking strategy.

Runs	Balasore District Dataset		Puri District Dataset		Cuttack District Dataset	
	Hypothesis Test	p-Value	Hypothesis Test	p-Value	Hypothesis Test	p-Value
	1	1	0.03316396	1	0.022158879	1
2	1	0.03426353	1	0.022165374	1	0.045182873
3	1	0.03326354	1	0.022263667	1	0.044762783
4	1	0.03387623	1	0.021773664	1	0.045002388
5	1	0.03316538	1	0.022377488	1	0.045288384

- Again, the performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Rabi and Kharif seasons crop prediction are done to obtain the minimum number

of features contributing towards rice crop yield and shown in **Table 9** and **Table 10**. From those tables, it can be concluded that, the features affecting mostly for rice yield are RH during 8.30 AM and 5.30



PM for all three districts taken during both the Rabi and Kharif season and also the minimum temperature plays a vital role.

- The paired T-test was used to calculate the importance of proposed fusion of feature ranking approach. The outcomes found by ELM-SVR-RFE were compared with proposed approach for five independent runs considering top five ranked features. Here, only ELM-SVR-RFE for statistical validation has been considered for paired test, as it gives healthier result than other basic feature ranking based methods.
- It can be observed from Table 11 and Table 12 that, the null hypothesis is rejected in case of Rabi season for all the three districts such as: Balasore, Puri and Cuttack and for three districts of Kharif season, as the values are closer to zero, which strengthens the argument that, proposed fusion of feature ranking approach has improved performance than the other only feature ranking based methods.

## 7 Conclusion and future scope

In this study an attempt has been made to obtain the climatic effect on rice yield of coastal areas of Odisha. The fusion based strategy is the novelty of this work. This prediction model not only predicts the rice yield per hectare but also able to obtain the significant or most affecting features during Rabi and Kharif seasons. This methodology works in three phases, in the first phase, three feature ranking approaches such as; Random Forest, SVR-RFE and F-Test has been applied on the three two datasets of three coastal areas and features are ranked as per their algorithm. In the second phase, Borda Count as a fusion method has been implemented on those ranked features from the above phase to obtain top five best features. Then in the third phase, multi-quadratic based ELM has been used to predict the rice crop yield using those ranked features obtained from fusion based ranking strategy of second phase. After applying ELM with fusion strategy, it is seen that by taking at least 3 features for Balasore, 5 features for Cuttack and 2 features for Puri we can get the accuracy of 99% where as in each individual ranking method with ELM we have to take more features. Finally, the statistical paired T-test has been used to evaluate and validate the significance of proposed fusion based ranking prediction model. From the observations made during experimentation, it has been found that; relative humidity and in some case temperature also is playing a vital role for rice crop production both for the Rabi season and the Kharif season. However, in future, the not linked or inconsequential factors can be later dealt with by working on optimized strategies.

## Acknowledgement

This work is financially supported by the Ministry of Science and Higher Education of the Russian Federation

(Government Order FENU-2020-0022).

## References

- [1] Central Soil and water Conservation Research & Training Institute (CSWCR & TI), Vision 2030, <http://www.cswcrtiweb.org/>. (Accessed on 17/10/2014).
- [2] Venkateswarlu, B. (2010). The 21st Dr. SP Raychaudhuri Memorial Lecture-Climate change: Adaptation and mitigation strategies in rainfed agriculture. *Journal of the Indian Society of Soil Science*, 58, S27-S35.
- [3] Saseendran, S. A., Singh, K. K., Rathore, L. S., Singh, S. V., & Sinha, S. K. (2000). Effects of climate change on rice production in the tropical humid climate of Kerala, India. *Climatic Change*, 44(4), 495-514.
- [4] Sarker, M. A. R., Alam, K., & Gow, J. (2012). Exploring the relationship between climate change and rice yield in Bangladesh: An analysis of time series data. *Agricultural Systems*, 112, 11-16.
- [5] Soora, N. K., Aggarwal, P. K., Saxena, R., Rani, S., Jain, S., & Chauhan, N. (2013). An assessment of regional vulnerability of rice to climate change in India. *Climatic Change*, 118(3-4), 683-699.
- [6] Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and electronics in agriculture*, 128, 67-76.
- [7] Gilbertson, J. K., & Van Niekerk, A. (2017). Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture*, 142, 50-58.
- [8] Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for Big Data analytics in plants. *Trends in plant science*, 19(12), 798-808.
- [9] Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, 140, 103-119.
- [10] Razmjoo, A., Xanthopoulos, P., & Zheng, Q. P. (2017). Online feature importance ranking based on sensitivity analysis. *Expert Systems with Applications*, 85, 397-406.
- [11] Teisseyre, P. (2016). Feature ranking for multi-label classification using Markov networks. *Neurocomputing*, 205, 439-454.

- [12] Lee, J., & Kim, D. W. (2015). Fast multi-label feature selection based on information-theoretic feature ranking. *Pattern Recognition*, 48(9), 2761-2771.
- [13] Fakhraei, S., Soltanian-Zadeh, H., & Fotouhi, F. (2014). Bias and stability of single variable classifiers for feature ranking and selection. *Expert systems with applications*, 41(15), 6945-6958.
- [14] Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, 15(6), 1437-1447.
- [15] Wei, C. C. (2013). Soft computing techniques in ensemble precipitation nowcast. *Applied Soft Computing*, 13(2), 793-805.
- [16] Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2017). META-DES. Oracle: Meta-learning and feature selection for dynamic ensemble selection. *Information fusion*, 38, 84-103.
- [17] Damiński, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., & Komorowski, J. (2008). Monte Carlo feature selection for supervised classification. *Bioinformatics*, 24(1), 110-117.
- [18] Tripoliti, E. E., Fotiadis, D. I., & Manis, G. (2013). Modifications of the construction and voting mechanisms of the random forests algorithm. *Data & Knowledge Engineering*, 87, 41-65.
- [19] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [20] Zhang, H. R., & Min, F. (2016). Three-way recommender systems based on random forests. *Knowledge-Based Systems*, 91, 275-286.
- [21] Wu, Q., Ye, Y., Zhang, H., Ng, M. K., & Ho, S. S. (2014). ForesTexter: an efficient random forest algorithm for imbalanced text categorization. *Knowledge-Based Systems*, 67, 105-116.
- [22] Yeh, C. C., Lin, F., & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33, 166-172.
- [23] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [24] Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353-363.
- [25] Shieh, M. D., & Yang, C. C. (2008). Multiclass SVM-RFE for product form feature selection. *Expert Systems with Applications*, 35(1-2), 531-541.
- [26] Mishra, S., & Mishra, D. (2015). SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm. *Karbala International Journal of Modern Science*, 1(2), 86-96.
- [27] Xu, Q., Kamel, M., & Salama, M. M. (2004, September). Significance test for feature subset selection on image recognition. In *International Conference Image Analysis and Recognition* (pp. 244-252). Springer, Berlin, Heidelberg.
- [28] Golugula, A., Lee, G., & Madabhushi, A. (2011, August). Evaluating feature selection strategies for high dimensional, small sample size datasets. In *2011 Annual International conference of the IEEE engineering in medicine and biology society* (pp. 949-952). IEEE.
- [29] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.
- [30] Das, S. R., Mishra, D., & Rout, M. (2019). A hybridized ELM using self-adaptive multi-population-based Jaya algorithm for currency exchange prediction: an empirical assessment. *Neural Computing and Applications*, 31(11), 7071-7094.
- [31] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H., & Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*, 27(1), 67-78.
- [32] Balasundaram, S., & Gupta, D. (2016). Knowledge-based extreme learning machines. *Neural Computing and Applications*, 27(6), 1629-1641.
- [33] Orissa Agricultural Statistics Year Book, (1983-2013). *Directorate of Agriculture and Food Production, Govt. of Odisha, Bhubaneswar*.
- [34] <https://www.google.co.in/images>
- [35] Narasimhamurthy, V., & Kumar, P. (2017). Rice Crop Yield Forecasting Using Random Forest Algorithm. *Int. J. Res. Appl. Sci. Eng. Technol. IJRASET*, 5, 1220-1225.
- [36] Dahal, H., & Routray, J. K. (2011). Identifying associations between soil and production variables using linear multiple regression models. *Journal of Agriculture and Environment*, 12, 27-37.
- [37] Powell, J. P., & Reinhard, S. (2016). Measuring the effects of extreme weather events on yields. *Weather and Climate extremes*, 12, 69-79.
- [38] Yusof, M. F., Azamathulla, H. M., & Abdullah, R. (2014). Prediction of soil erodibility factor for Peninsular Malaysia soil series using ANN. *Neural Computing and Applications*, 24(2), 383-389.

- [39] Erdil, A., & Arcaklioglu, E. (2013). The prediction of meteorological variables using artificial neural network. *Neural Computing and Applications*, 22(7-8), 1677-1683.
- [40] Anitha, A., & Acharjya, D. P. (2018). Crop suitability prediction in Vellore District using rough set on fuzzy approximation space and neural network. *Neural Computing and Applications*, 30(12), 3633-3650.
- [41] Zahid, M. A., & De Swart, H. (2015). The borda majority count. *Information Sciences*, 295, 429-440.
- [42] García-Lapresta, J. L., Martínez-Panero, M., & Meneses, L. C. (2009). Defining the Borda count in a linguistic decision making context. *Information Sciences*, 179(14), 2309-2316.
- [43] <https://www.casact.org/pubs/forum/98wforum/98wf055.pdf>
- [44] Hirai, GI., Chiyo, H., Tanka, O., Hikano, T., & Oanotri, M. (1993). Studies on the effect of relative humidity of atmosphere on growth and physiology of rice plants. VIII effect of ambient humidity on dry matter production and nitrogen absorption at various temperatures, *Japanese Journal of Crop Science*, 62(3), 395-400.
- [45] Sunil, K. M. (2000). Crops weather relationship in rice (*Doctoral dissertation, Department of Agricultural Meteorology, College of Horticulture, Vellankkara*).
- [46] Vijayakumar, CM. (1996). Hybrid rice seed production technology- theory and practice. *Directorate of rice research, Hyderabad*, 52-55.
- [47] Gridyal, B. P., & Jana, R. K. (1997). Agrometeorological environmental affecting rice yield. *Agronomy Journal*, 59, 286-287.
- [48] Narayanan, A. L. (2004). Relative influence of weather parameters on rice hybrid and variety and validation of CERES- Rice model for staggered weeks of transplanting. *PhD Thesis, Tamilnadu Agricultural University, Coimbatore*.
- [49] Shi, C. H., & Shen, Z. T. (1990). Effect of high humidity and low temperature on spikelet fertility in indica rice. *International Rice Research Newsletter*, 15(3), 10-11.
- [50] Morita, S., Wada, H., & Matsue, Y. (2016). Countermeasures for heat damage in rice grain quality under climate change. *Plant Production Science*, 19(1), 1-11.

