A Classifier Ensemble Approach for Prediction of Rice Yield Based on Climatic Variability for Coastal Odisha Region of India

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: crop prediction, classifier ensemble, support vector machine, k-nearest neighbour, naive bayesian, decision tree, linear discriminant analysis

Received: February 23, 2021

Agriculture is the backbone of Indian economy especially rice production, but due to several reasons the expected rice yields are not produced. The rice production mainly depends on climatic parameters such as rainfall, temperature, humidity, wind speed etc. If the farmers can get the timely advice on variation of climatic condition, they can take appropriate action to increase the rice production. This factor motivate us to prepare a computational model for the farmers and ultimately to the society also. The main contribution of this work is to present a classifier ensemble based prediction model by considering the original rice yield and climatic datasets of coastal districts Odisha namely Balasore, Cuttack and Puri for the period of 1983 to 2014 for Rabi and Kharif seasons. This ensemble method uses five diversified classifiers such as Support Vector Machine, k-Nearest Neighbour, Naive Bayesian, Decision Tree, and Linear Discriminant Analysis. This is an iterative approach; where at each iteration one classifier acts as main classifier and other four classifiers are used as base classifiers whose output has been considered after taking the majority voting. The performance measure increases 95.38% to 98.10% and 95.38% to 98.10% for specificity, 88.48% to 96.25% and 83.60% to 94.81% for both sensitivity and precision and 91.78% to 97.17% and 74.48% to 88.59% for AUC for Rabi and Kharif seasons dataset of Balasore district and also same improvement in Puri and Cuttack District. Thus the average classification accuracy is found to be above 96%.

Povzetek: Opisana je ansambelska metoda napovedovanja pridelka riža v Indiji.

1 Introduction

Agriculture is the pivot of Indian economy. Around 58% of rural households are dependent on agriculture as their major means of livelihood. However, the share of agriculture has changed considerably in the past 50 years. In 1950 55% of GDP came from agriculture while in 2009 it is 18.5% and during the financial year 2015-2016 it is 16.85% [1]. Indian agriculture has made great progress in ensuring food security to its huge population with its food grains production reaching a record level of 236 million ton in 2013-2014. While the required amount for 2030 and 2050 are 345 and 494 million ton respectively. In India rice is grown in different agro climatic zones and altitudes. Rice grown in India has extended from 8 to 35°N latitude and from sea level to 3000 meter. Rice required a hot and humid climate and well suited to the areas which have high humidity, long

sun shine and sufficient water supply. The average temperature required for the crop is 21 to 36°C. It is predicted that the demand for the rice will grow further than other crops. There are various challenges to achieve higher productivity with respect to climate change and its repercussions. In tropical area higher temperature is one of the important environmental factors which limit rice production. Different parts of the country have variable impacts due to climate change. For example by the year of 2080 the numbers of rain days are to be decreased along with narrow rise of 7-10% annual rain fall which will lead to high intensity storm. Moreover, on one hand when monsoon rain fall over the country is expected to rise by 10-15%, on the other hand the winter rain fall is expected to reduce by 5-26% and seasonal variability would be further compounded [2]. Then, cereal production is expected to be reduced by 10-40% by 2100 due to rise in temperature, rising water scarcity and decrease in number of rain days. Higher loss is predicted in *Rabi* crops [3]. Rice productivity may decline by 6 percent for every 10C rising temperature [4]. In general changing climate trends will lead to overall decline agricultural yield. The simulation analysis projected that on all India basis, the consequent of climate change on productivity in 2030s ranges from -2.5 to -12% for crops such as rice, wheat, maize, sorghum, mustard and potato [5, 6]. Climate is the sum of total variation in temperature, humidity, rainfall and other metrological factors in a particular area for a period of at least 25 years [1]. Odisha's climate has also under gone appreciable changes as a result of various factors. The previous six seasons of the year has changed into basically two mainly summer and rainy. The deviation in day temperature and annual precipitation is mainly restricted to 4 months in a year and number of rain days decreased from 120 to 90 days apart from being abnormal. In addition, the mean temperature is increasing and minimum temperature has increased about 25% [2, 3, 4, 5]. Such climate change related adversity is affecting adversely productivity and production of food grains. Agriculture is the backbone of Indian economy. But due to several reasons the expected crop yields are not produced. The production mainly depends on climatic parameters such as rainfall, temperature, humidity, wind speed etc. So the farmer should know the timely variation in climatic condition. If they can get the timely advice then they can increase the production. Before development of the technology the farmers can predict the production just by seeing the previous experience on a particular crop. But gradually the data increases and due to the environmental factors the weather changes. So we can use this vast amount of data for prediction of rice production. For a uniform growth and development assurance in agriculture (the current rate is 2.8% per annum), an exhaustive appraisal of the accountability of the agriculture production owing to predicted type of weather transform is necessary.In this paper the main aim is to create an ensemble model for prediction of climatic variability on rice yield for coastal Odisha. The weather parameters such as rainfall, temperature and humidity etc. are considered because they affect the 95% production of rice crop. Additionally, the classifier's accuracy validity has been measured using specificity, sensitivity/recall, precision, Negative Predictive Value (NPV), False Positive Rate (FPR), False Negative Rate (FNP), False Discovery Rate (FDR) and the probabilistic measures such as; F-Score, G-Mean, Matthews Correlation Coefficient (MCC) and J-Statistics. This paper is organized as follows; section 2 describes the related works, materials and methods or approaches used for experimentation are described in section 3. The framework of the proposed prediction model is given in section 4, section 5 deals with experimentation and model evaluation. The result analysis, discussion and conclusion are given in section 6, 7 and 8 respectively.

2 Related work

While undertaking this work, the existing literature that has been followed during every phase of the entire research work with the intention of clear representation of the machine learning based prediction models. The various approaches are explored and have been addressed to design the ensemble based rice production model based on climatic variability. This section describes few recent works on this are which motivated us to develop an ensemble based model. Narayan Balkrishnan [7] proposed an ensemble model AdaSVM and AdaNaive which is used to project the crop production. Authors compared their proposed model among the Support Vector Machine (SVM) and Naïve Bayes (NB) methods. For prediction of output, two parameters are used such as accuracy and the classification error and it has been observed that AdaSVM and AdaNaive are better than SVM and NB. B Narayanan [8][8] compared the SVM and NB with AdaSVM and AdaNaive and conclude that the later one is better than first two methods. Sadegh Bafandeh [9] studied the detailed historical background and different applications of the method in various areas. If the distribution of the data is not known then the k-Nearest Neighbour (K-NN) method can be applied for classification technique [10, 11, 12]. In the feature space objects can be classified on the basis of closest training examples. It is one of the instance–based learning or lazy learning where computation is done until classification and function is approximated locally [13, 14]. A Bayesian network or Bayes network or belief network or Bayesian model or probabilistic directed acyclic graphical models a type of statistical model. A belief network to assess the effect of climate change on potato production was formulated by yiqun Gu et. al. [15]. Authors have shown a belief network combining the uncertainty of future climate change, considering the variability of current weather parameters such as temperature, radiation, rainfall and the knowledge about potato development. They thought that their network give support for policy makers in agriculture. They test their model by using synthetic weather scenarios and then the results are compared with the conventional math-

ematical model and conclude that the efficiency is more for the belief network. There are various factors influencing the prediction. UnoY et al. [16] used agronomic variables, nitrogen application and weed control using the machine learning algorithm such as artificial neural network and Decision Tree (DT) to develop the yield mapping and to forecast yield. They have concluded that high prediction accuracies are obtained by using ANNs. Veenadhari S et al. [17] described the soybean productivity modelling using DT algorithms. Authors have collected the climate data of Bhopal district for the period 1984-2003. They considered the climatic factors such as evaporation, maximum temperature, maximum relative humidity, rainfall and the crop was soybean yield and applied the Interactive Dichotomizer3 algorithm which is information based method and based on two assumptions. Using the induction tree analysis it was found that the relative humidity is a major influencing parameter on the soybean crop yield. DT formed for influence of climatic factors on soybean yield. Using the if-then-else rules the DT is formulated to classification rules. Relative humidity affects much on the production of soybean and some rules generated which help to in the low and high prediction of soybean. One of the drawbacks was only the low or high yield can be predicted but the amount of yield production cannot be predicted. Due to the diversity of climate in India, agriculture crops are poorly impressed in terms of their achievement from past two decades. Forecasting of crop production and advanced yield might be helpful to policy inventor and farmers to take convenient decision. The forecasting also helps for planning in the industries and they can coordinate their business on account of the component of the climate. A software tool titled 'Crop Advisor' has been developed by Veenadhari et al. [18] which is a client friendly and can forecast the crop yields with the effect of weather parameters.C4.5 algorithm is applied ascertain the most effective climatic parameter on the crop yields of specified crops in preferred district of Madhya Pradesh. The software will be helpful for advice the effect of various weather parameters on the crop yield. Other agro –input parameters liable for crop yield are not accommodating in this tool, since the application of these input parameters differ with individual fields in space and time. Alexander Brenning et al. [19] compared all the classifier including Linear Discriminant Analysis (LDA) for crop identification based on multitemporal land dataset and concluded that stabilized LDA performed well mainly in field wise classification. Minggang Du et al. [20] used the method LDA for plant classification and conclude that LDA with Principal Component Analysis is effective and feasible for plant classification. Renrang Liao [21] classified fruit tree crops using penalized LDA and found that the LDA may not be able to deal with collinear high dimensional data. It has been observed that, most of literature are using single classification model to predict the crop yield leading to increase in misclassification by data biasing, therefore we have been motivated to formulate a multiclassifier based model known as classifier ensemble [22]. This ensemble technique helps to reduce the classification error by considering the outputs of different classifiers by taking the majority of right outputs [23, 24]. In this paper we have tried to consider the collision of the weather transform scenario of Odisha context of the farming yield of the one main fasten food rice using the machine learning methods such as SVM, *K*-NN, NB, DT and LDA [25, 26].

3 Materials and methods

This section briefly describes the machine learning techniques and tools used to develop the ensemble based crop prediction model.

3.1 Support Vector Machine (SVM)

TSVM is one of the supervised machine learning techniques and also known as support vector networks. It analyses data mainly for classification and regression analysis. A set of labelled training data it produces by using input-output mapping functions [27]. For both classification of linear and non linear dataset, SVM method can be used. The original training data transformed a higher dimension by SVM using non linear mapping. Then for the linear optimal separating hyper plane, the new dimension searched by SVM. Thus, a decision boundary formed which separates the different classes from one another [28]. When the SVM is used for the prediction of the crop yield then it is known as support vector regression. The main objective of the SVM is to find non-linear function by the use of kernel that is a linear on polynomial function [29, 30, 30]. The radial basis function and the polynomial function are the widely used kernel functions. In case large input samples space the difficulty of using linear function can be avoided by using SVM. Due to optimization the complex problem can be converted into simple linear function optimization [32].

3.2 *K*-Nearest Neighbour (*K*-NN)

K-NN [33] is one of the simplest supervised learning methods used for both classification and prediction techniques [34, 35]. By using *K*-NN the unknown sample can be classified to predefined classes, based on the training data. It requires more computation than other techniques. But it is better for dynamic numbers that change or updated quickly. For new sample classification the *K*-NN process the detachment among the entire sample in the training data. The Euclidian distance is used for distance measurement. The samples with the smallest distance to the new sample are known as *K*-nearest neighbours [36]. The main idea behind the*K*-NN is to estimate on a fixed number of observations those are closest to the desired output. It can be used for both in discrete and continuous decision making such as classification and regression. In case of classification most frequent neighbours are selected and in case

of prediction or regression the average of k-neighbours are calculated. Besides the Euclidean distance, Manhattan distance and Minkowski distance are used in *K*-NN [37].

3.3 Naïve Bayesian Classifier (NB)

The NB classification technique is developed on the basis of Bayesian theorem. This technique is most suitable when the input value is very high that when the dataset is very high we can use the Naïve Bayes technique. The other names of Bayes classifiers are simple Bayes or idiot Bayes [38]. Naïve Bayes classifier is a simple probabilistic classifier with strong independence assumptions. The classifier can be trained on the nature of the probability model. It can work well in many complex real world situations. It requires a little quantity of training data to calculate the parameter essential for the classification and it is the main advantage of Naïve Bayes classifier. Bayes theorem is based on probabilistic belief. It is based on conditional probability on mathematical manipulation. Therefore, Bayes important characteristics can be computed using rules of probability, more specific conditional probability [39].

3.4 Decision Tree (DT)

DT presents a very encouraging technique for automating most of the data mining and predictive modelling process. They embed automated solutions such as over fitting and handling missing data. The models built by DTs can be easily viewed as a tree of simple decisions and provide well-integrated solutions with high accuracy. DT also known as classification tree is a tree like structure which recursively partitions the dataset in terms of its features. Each interior node of such a tree is labelled with a test function. The best known DT algorithms are C4.5 and ID3 [40].The figure 1 illustrates an example of DT with their IF ... THEN... ELSE ... rules form.

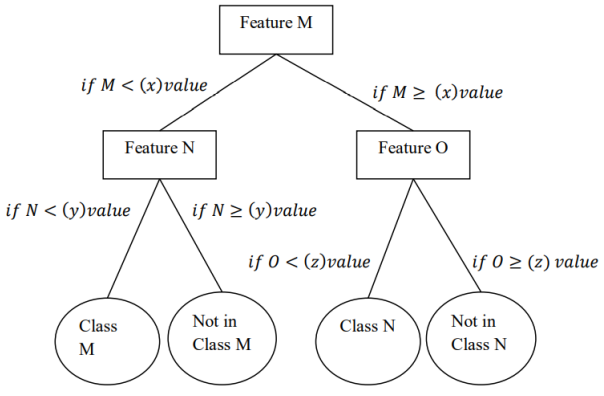


Figure 1: Decision Tree with IF ... THEN ... ELSE ... Rules form

3.5 Linear Discriminant Analysis (LDA)

Discriminant analysis is a multivariate method of classification. Discriminant analysis is similar to regression analysis except that the dependent variable is categorical rather than continuous in discriminant analysis; the intent is to predict class membership of individual observations based on a set of predictor variables. LDA generally attempts to find linear combinations of predictor variables that best separate the groups of observations. These combinations are called discriminant functions. It is one of the dimensional reduction methods, used in preprocessing in pattern-classification and machine learning applications. In order to avoid over fitting we can apply LDA in the dataset for good class separability with reduced computational cost [41]. Linear combinations of the predictors are used by LDA to model the degree to which an observation belongs to each class and discriminant function is used and a threshold is applied for classification [42].

3.6 Majority voting

Majority voting is one of the ensemble learning algorithms, which is a voting based methods. Majority vote is appropriate when each classifier cl can produce class-probability estimates rather than a simple classification decision. *A class-probability estimate* for data point y is the probability that the true class is $k : A(f(x) = m|cl)$, for $m = 1, \cdots, M$. We can combine the class probabilities of all the hypotheses so that the class probability of the ensemble can be found [43]. Sarwesh Site et. al. described about the better performance for better prediction after merging two or more classifier using the voting of data, which is known as ensemble classifier. They described various technique of ensemble classifier both for binary classification and multi-class classification [44]. Xueyi Wang et. al. prepared a model to find the accuracies of majority voting ensembles by taking the UCI repository data and made experiment of the 32 dataset. They made their data into different subsets such as core, outlier and boundary and found result that for better ensemble method or to achieve high accuracy; the weak individual classifier should be partly diverse [45].

3.7 Performance measures

This section discusses the basics of *specificity*, *sensitivity/recall*, and *precision*, NPV, FPR, FNP, FDR, F-Score, G-Mean, MCC and J-Statistics. These are extent to which a test measures what it is supposed to measure; in other words, it is the accuracy of the test or validity of the test and measured using a *confusion matrix* i.e. a two-by-two matrix. There are four elements of a confusion matrix such as; *True Positives (TP)*, *False Positives (FP)*, *False Negatives (FN)* and *True Negatives (TN)* represented in the a, b, c and d cells in the matrix []. *Specificity* is computed as $d(TN)/(FP) + d(TN)$, *sensitivity* as; $a(TP)/a(TP) + c(FN)$. *Sensitivity* and *specificity* are inversely proportional, i.e. as the sensitivity increases, the

specificity decreases and vice versa. *Precision* tells about, how many of test positives are true positives and if this number is higher or closer to 100 then, this test it suggests that this new test is doing as good as the defined standard. It can be computed as; $a(TP)/a(TP)+b(FP)$; NPV tells how many of test negatives are true negatives and the desired value is approximately 100 and then it suggests that this new test is doing as good as the defined standard. Computed as; $d(TN)/c(FN)+d(TN)$. Assuming all other factors remain constant, the PPV will increase with increasing prevalence; and NPV decreases with increase in prevalence.A false positive error or fall-out is a result that indicates a given condition has been fulfilled, when it actually has not been fulfilled, or erroneously a positive effect has been assumed. In other words, it is the proportion of all negatives that still yield positive test outcomes, i.e., the conditional probability of a positive test result given an event that was not present and computed as $b(FP)/b(FP)+d(TN)$ or 1-*Specificity*. An FNR is a test that result indicates a condition failed, while it actually was successful, or erroneously no effect has been assumed. In other words, it is the proportion of events that are being tested for which yield negative test outcomes with the test, i.e., the conditional probability of a negative test result given that the event being looked for has taken place and can be computed as, $c(FN)/a(TP)+c(FN)$ or 1-*Sensitivity*. FDR is a way of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons. FDR-controlling procedures are designed to control the expected proportion of rejected null hypotheses that were incorrect rejections or false discoveries and computed as, $b(FP)/b(FP)+a(TP)$ or 1-PPV. F-Score measure considers both the precision and the recall of the test to compute the score. It can be interpreted as a weighted average of the precision and recall, where an F-Score reaches its best value at 1 and worst at 0. It can be computed as: $2\times((Precision\times Recall)/(Precision\times Recall))$. MCC is used to measure the quality of binary classification. It takes into account true and false positives and negatives and is generally regarded as a balanced measure and can be used in case of imbalanced datasets. This is a correlation coefficient between the observed and predicted binary classification results. While there is no perfect way of describing the confusion matrix of true and false positives and negatives by a single number, MCC is generally regarded as being one of the best such measures and can be computed as: $((a\times d)-(b\times c))/\sqrt{(a+b)\times(a+c)\times(d+b)\times(d+c)}$. The accuracy determined for the classifiers may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases i.e. the imbalanced classes. Suppose, there are 1000 cases, 995 of which are negative cases and 5 of which are positive cases. If the system classifies them all as negative, the accuracy would be 99.5% even though the classifier missed all positive cases, in such cases G-mean comes into action. G-mean has the maximum value when sensitivity and specificity are equal and can be computed as: $\sqrt{Precision\times Recall}$. Youden's J Statistics is a way of summarizing the performance of a diagnostic test. For a test with poor diagnostic accuracy, Youden's index equals 0, and in a perfect test Youden's index equals 1. The index gives equal weight to false positive and false negative values, so all tests with the same value of the index give the same proportion of total misclassified results. This is $Sensitivity+Specificity\breve{}1$.

4 Structural and functional representation of proposed ensemble based prediction model

The schematic representation of the proposed model is shown in Figure 2. First the datasets are collected from three coastal district of Odisha and different parameters collected from the Odisha Agriculture Statistics, Director of Agriculture and Food Production, Govt. of Odisha, Bhubaneswar sources, and then the datasets are pre-processed. The proposed methodology is based on classifier ensemble method. The intension is to predict the rice yield for two seasons such as Rabi and Kharif with respect to the climatic variability of the coastal Odisha. This model uses five classifiers where four classifiers act as base classifier and one act as main classifier. List of classifier used are SVM, k-NN, DT, NB and LDA. Experiments are conducted by considering each classifier once as main classifier and remaining four as base classifiers by using MATLAB 10 at windows OS. Then, we get five different predicted outputs for rice production. Each classifier is build according the basic algorithm defined in literature [26] [31] [36] [38] [40] [43].

Let $B=\{b_1,\cdots,b_4\}$ be the four base classifiers, and $C=\{c_1,\cdots,c_4\}$ be the output of those four base classifiers. The output of each classifier is passed through a conversion function f to retrieve the production denoted as $\hat{S}$ as given below and this acts as input to main classifier.

$$\hat{S}_l=f(c_i) \tag{1}$$

Where f can be computed using equation (2)

$$f(c_i)=\frac{N}{|N|} \tag{2}$$

Where N is the sum of S_i which belongs to class c_i

Hence, main classifier will have input having *vector* $D=\{dataset,\hat{S}_1,\hat{S}_2,\hat{S}3,\hat{S}_4\}$. Result obtained after processing D by main classifier is compared expected output (y). Again equation (2) is used to compute the production based upon the class labels predicted.

Final prediction is made by using majority voting on the class label predicted by each classifier as main classifier (Figure 3). Throughout the paper **#** symbol is used before classifier for differentiating it with base classifiers

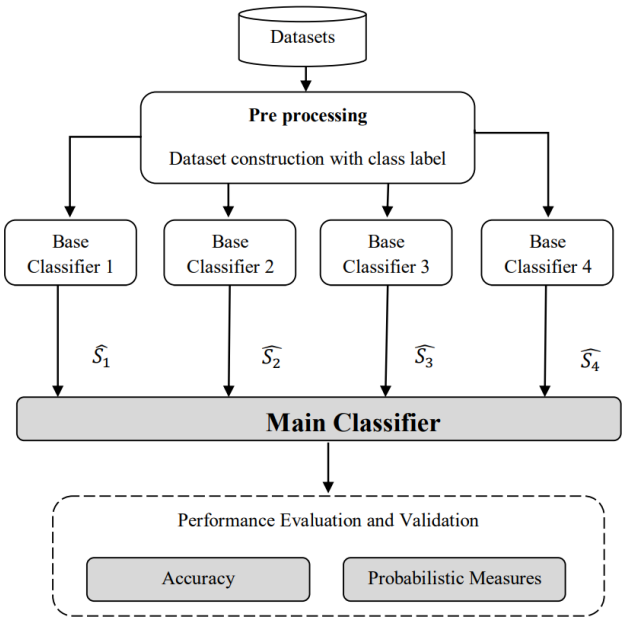


Figure 2: Schematic representation of proposed ensemble based prediction model

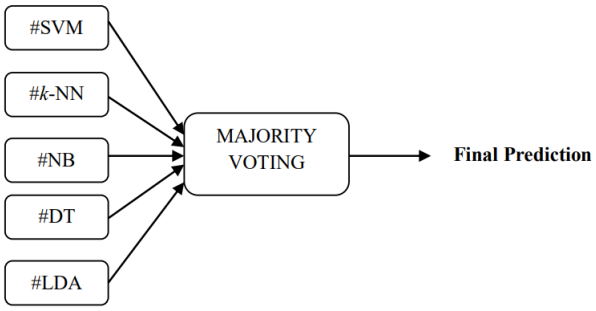


Figure 3: Majority voting applied on the main classifiers

5 Experimentation and model evaluation

This section elaborates the experimentation process starting from datasets chosen with their description, step wise representation of the working principle of proposed method and also the results are analyzed with respect to the average classification accuracy and the predictive performances used to validate the model.

5.1 Dataset description

Real dataset D is collected from three coastal regions of Odisha such as Balasore, Puri, Cuttack district. Let $d_i \in D \ \forall i = 1, \cdots, 31$ features where $|d_i| = 25$ represents the attributes of the datasets. Different parameters collected from the Odisha Agriculture Statistics, Director of Agriculture and Food Production, Govt. of Odisha, Bhubaneswar [46]; such as $p = \{max\ temperature, min\ temperature, rain\ fall, humidity\}$ that effect the rice production. Since, there are two types of rice production seasons such as; *Rabi* and *Kharif* produced between months "January - June" and "July –December", hence p_i is collected over the range of six months each resulting 24 set of attributes and 25$^{\text{th}}$ 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 (4a) and Figure (4b) for *Rabi* and *Kharif* season respectively. The detail description of datasets with standard deviation (Std. Dev) for three areas is shown in Table 1.

5.2 Construction of dataset for classification

Raw data collected have some missing value, and without class. One way is to deal with missing value is to simply replace it with most negligible positive real number. For classification, D must be in the form $D = \{d, y\}$, where d_i refers to *features* and y_i refers to *class label*. In order to predict the production of rice crop, one needs to properly define class label. One way is to use clustering and allocate each feature a class label similar to their cluster number. Looking to the random cluster index formed makes it difficult to build common class label for the feature. Hence, in our work we have proposed a *range based class label formation*. Let Sdenote the production column vector of dataset D and y_i can be formulated using equation (3).

$$y_i = \begin{cases} u \leq s_i < r & 1 \\ r \leq s_i < 2 \times r & 2 \\ r \leq s_i < 3 \times r & 3 \\ \cdots \quad \cdots & \cdots \\ k \times r \leq s_i < v & k \end{cases} \tag{3}$$

Where, $[u, v]$ is the min and max value of S given by equation (4), r is the offset for range formation given by equation (5) and $k = 5$. **Table 2** shows the number of year

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

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

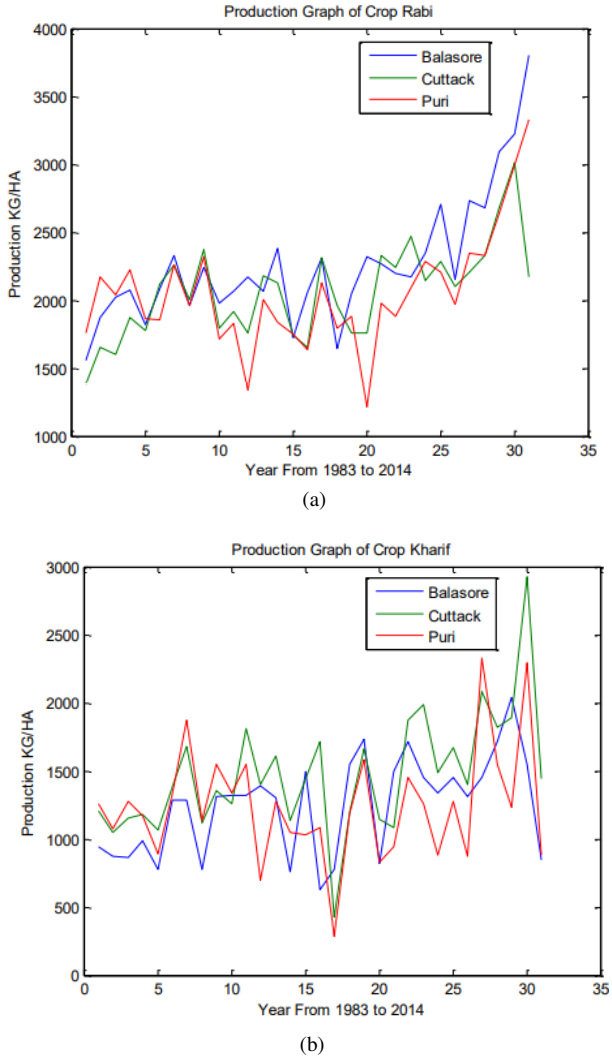


Figure 4: Graphical representation of rice production of three regions for *Rabi* and *Kharif* seasons

belonging to different k classes. That means, the total data of 31 years is divided into 5 classes.

$$u = min(S), \quad v = max(S) \tag{4}$$

$$r = (u - v)/k \tag{5}$$

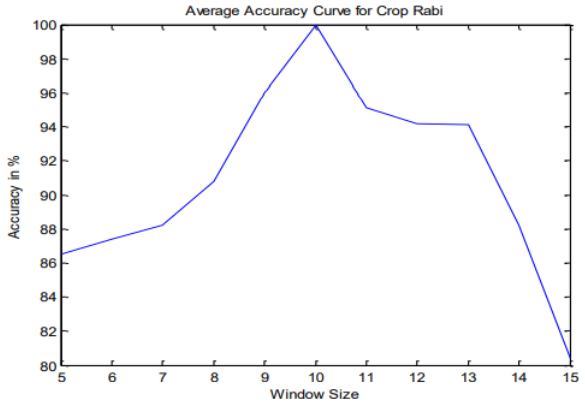
6 Result and performance analysis

Proposed architecture is implemented on Matlab 10 at Windows OS with min 2GB RAM and 2 GH Intel Processor. Dataset is given as input to the proposed architecture using sliding window concept. Window size of w feature is used for training and feature $w + 1$ is used for testing. **Figure (5a)** and (**5b**) shows the average accuracy curve gained by different set of window sizes w for *Rabi* and *Kharif* season crop productions respectively. From the both the figures it can be observed that for the window size of 10 and 12 the proposed architecture accuracy reaches 100% for *Rabi* and *Kharif* season datasets respectively.

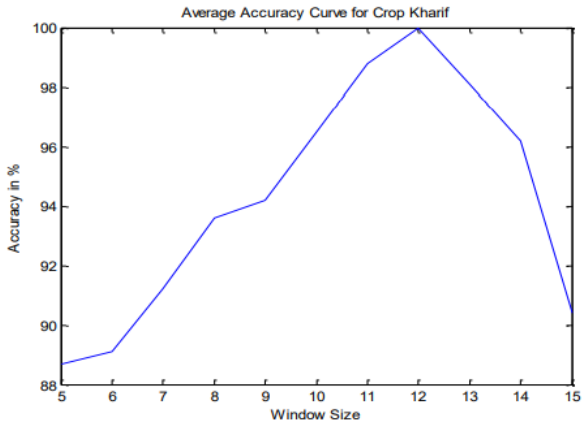
During the literature survey, we have explored various methods already used and found that the ensemble methods give better result in most of the cases. Then we have analysed all the ensemble methods and consider SVM, *K*-NN, NB, DT and LDA classifiers for our experimentation. At each iteration; four classifiers are chosen as base classifiers and the output of those base classifiers ($\hat{S}$) are passed though the conversion function f as given in equation (1) and (2) to the main classifier. The main classifier containing the input vector $D = \{dataset, \hat{S}_1, \hat{S}_2, \hat{S}_3, \hat{S}_4\}$, does the prediction. The result obtained after processing D by main classifier is compared expected output y. Final prediction is made by using *majority voting* on the class label

Table 2: Class label determination according to $k = 5$

Datasets/ Seasons	**District**	**Class**				
		1	**2**	**3**	**4**	**5**
Rabi	Balasore	8	12	4	3	4
	Cuttack	4	8	13	3	3
	Puri	3	12	10	3	3
Kharif	Balasore	9	3	9	7	3
	Cuttack	3	11	12	2	3
	Puri	3	8	12	5	3



(a)



(b)

Figure 5: Accuracy curve for selection of window size (w) for training data (**a**) *Rabi* and (**b**) *Kharif* seasons

predicted by each classifier as main classifier after each iteration. This process has been implemented by considering the window size $w = 10$ and $w = 12$ for both the *Rabi* and *Kharif* seasons datasets respectively. The average accuracy obtained for prediction of rice production in hector for *Rabi* season in hectors is shown in **Table 3**. The prediction curve of rice for *Rabi* season dataset for Balasore, Cuttack and Puri is shown in **Figure (6a), (6b)** and **(6c)**. From the **Figure 7** we can see that, the MV line touches the actual value of production line more than other classifiers and it proves that the ensemble MV method is better than the individual classifier.

It is clear from the **Table 3** that if we are applying each individual four classifier such as SVM, K-NN, NB, DT and LDA as main classifier then with majority voting (MV) then the accuracy of MV gives better accuracy. In case of Balasore, MV gives 98.21% accuracy than the other classifiers. Similarly, the same improved performance in case of Cuttack and Puri district also. From the figure 7 we can see that, the MV line touches the actual value of production line more than other classifiers and it proves that the ensemble MV method is better than the individual classifier.

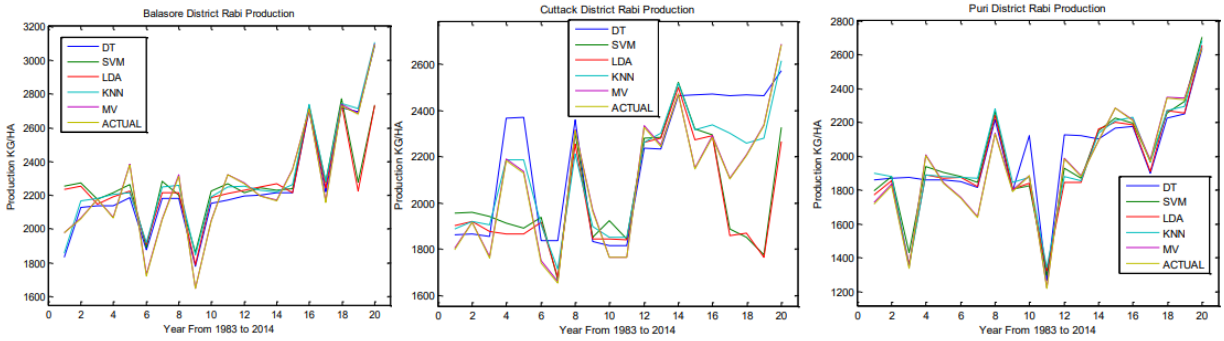
The average accuracy obtained for prediction of rice production in hector for Kharif season is shown from **Table 4**. The prediction curve of rice for Kharif season dataset for Balasore, Cuttack and Puri is shown in **Figure (7a), (7b)** and **(7c)**. In the **Table 4**, it shows that as in case of *Kharif* season dataset, the MV in the ensemble classifier gives better accuracy exceeding 96% for all three districts such as: Balasore, Cuttack, Puri like *Rabi* season. **Figure 7** shows that the MV line is touching the actual data line and gives the better result.

The datasets are imbalanced in nature i.e. the distribution of data elements for the classes varies a large giving rise to biased opinion and over generalization of classifiers towards a single class having large elements. In such type of situations, the average classification accuracy is not enough to prove the stability and validity of the classifiers. Therefore, in this paper, we have tried to establish the performance of proposed model by considering the specificity, sensitivity/recall, precision, NPV, FPR, FDR, F-Score, G-Mean, MCC and J-Statistics, and AUC. The value of each measure should lie between $[0 - 1]$, where 0 represents lower prediction ability and 1 represents the high prediction ability. The performance of the proposed prediction model for all three districts such as Balasore, Cuttack and Puri for *Rabi* season datasets are shown from **Table 5** to **Table 7** and from **Table 8** to **Table 10** for *Khariff* season datasets.

In the **Table 5** it shows that, the improvements of performance measures approaches towards 95.09% to 98.10% for specificity, 88.48% to 96.25% for both sensitivity and precision and 91.78% to 97.17% for AUC for *Rabi* season dataset of Balasore district. So we can see comparing all other main classifier, when SVM choosen as main classifier it gives better performance. Similarly for other performance measure the result is also like specificity.

Table 3: Average classification accuracy (%) of each classifier and one classifier as main classifier (preceded with **#**) for prediction of rice production in hector for *Rabi* season dataset

District	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA	MV
Balasore	86.29	80.61	82.25	84.88	81.25	97.48	95.91	93.11	95.62	93.40	98.21
Cuttack	87.99	84.06	86.48	86.45	85.93	95.67	94.55	94.08	96.79	95.16	97.13
Puri	89.61	88.99	87.60	90.39	92.02	99.25	96.11	95.83	98.33	94.60	99.61

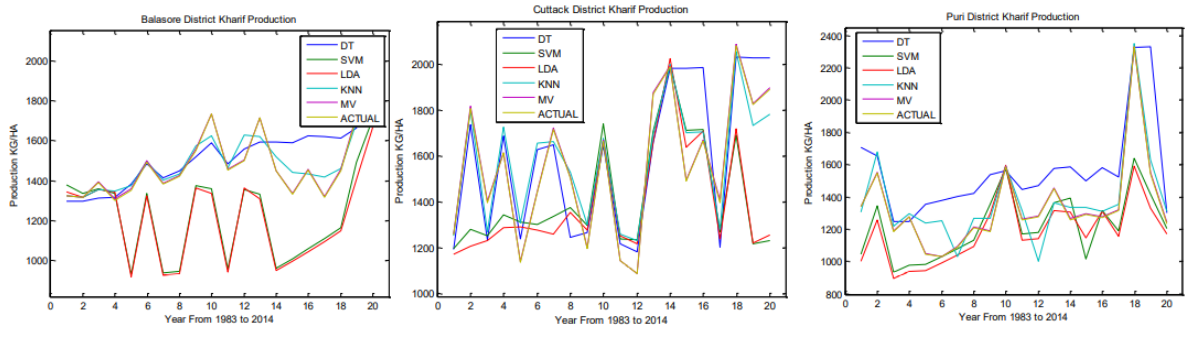


(a) (b) (c)

Figure 6: Rice production prediction curve for *Rabi* season dataset at (**a**) Balasore, (**b**) Cuttack, and (**c**) Puri districts

Table 4: Average classification accuracy (%) of each classifier and one classifier as main classifier (preceded with **#**) for prediction of rice production in hector for *Kharif* season dataset

District	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA	MV
Balasore	79.41	67.00	69.96	75.05	65.77	96.46	94.05	89.34	93.38	89.49	97.82
Cuttack	81.84	73.50	78.15	77.62	76.03	93.73	91.84	90.94	95.24	92.51	96.12
Puri	86.37	85.19	82.68	86.96	89.28	99.08	94.95	94.50	97.85	92.55	99.21



(a) (b) (c)

Figure 7: Rice production prediction curve for *Kharif* season dataset at (**a**) Balasore, (**b**) Cuttack, and (**c**) Puri districts

Table 6 shows the performance measure of *Rabi* season of Cuttack district. It seen that, the improvements of performance measures approaches towards 96.18% to 97.14% in case of specificity and NPV, 86.85% to 93.14% in case of sensitivity and precision. In this case when DT choosen as mainclassifier gives better result and in other performance measure also the same case. From the **Table 7** it can be seen that, the performance value improves from 96.80% to 99.54% in case of specificity, 82.55% to 98.02% in case of sensitivity and precision, 89.68% to 98.78% in case of AUC. Similarly others can be seen. In all case the SVM main classifier gives better result for Rabi season of Puri district.

In the **Table 8**, it is clear that, in case of *Kharif* season of Balasore district, SVM main classifier gives better performance than others in all performance measures.

In **Table 9** also the main classifier gives better performance than the individual classifier and the performance of specificity improves from 96.18% to 97.90%, 85.69% to 92.57% for sensitivity, precision and NPV. Here DT main

Table 5: Performance measures of rice production for *Rabi* season at Balasore district

Measures	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA
Specificity	89.83	86.39	87.57	89.35	87.27	98.10	96.99	95.09	96.85	95.38
Sensitivity	78.99	66.29	69.00	73.95	64.41	96.25	93.62	88.48	92.80	88.48
Precision	78.99	66.29	69.00	73.95	64.41	96.25	93.62	88.48	92.80	88.48
NPV	89.83	86.39	87.57	89.35	87.27	98.10	96.99	95.09	96.85	95.38
FPR	10.17	13.61	12.43	10.65	12.73	1.90	3.01	4.91	3.15	4.62
FNR	21.01	33.71	31.00	26.05	35.59	3.75	6.38	11.52	7.20	11.52
FDR5	21.01	33.71	31.00	26.05	35.59	3.75	6.38	11.52	7.20	11.52
F-Score	70.12	57.60	60.27	65.14	55.76	87.19	84.59	79.49	83.77	79.49
G-Mean	68.26	55.39	58.15	63.16	53.49	85.67	83.02	77.84	82.19	77.84
MCC	68.82	52.67	56.57	63.30	51.69	94.35	90.62	83.56	89.65	83.86
J-Statistics	65.76	49.84	53.68	60.31	48.90	91.13	87.41	80.39	86.45	80.68
AUC	84.41	76.34	78.28	81.65	75.84	97.17	95.31	91.78	94.83	91.93

Table 6: Performance measures of rice production for *Rabi* season at Cuttack district

Measures	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA
Specificity	92.18	89.96	91.42	91.46	91.22	97.14	96.45	96.18	97.9	96.89
Sensitivity	74.07	61.30	68.19	67.22	64.63	91.08	88.25	86.85	93.14	89.07
Precision	74.07	61.30	68.19	67.22	64.63	91.08	88.25	86.85	93.14	89.07
NPV	92.18	89.96	91.42	91.46	91.22	97.14	96.45	96.18	97.90	96.89
FPR	7.82	10.04	8.58	8.54	8.78	2.86	3.55	3.82	2.10	3.11
FNR	25.93	38.70	31.81	32.78	35.37	8.92	11.75	13.15	6.86	10.93
FDR	25.93	38.70	31.81	32.78	35.37	8.92	11.75	13.15	6.86	10.93
F-Score	65.26	52.71	59.46	58.51	55.97	82.07	79.26	77.88	84.11	80.08
G-Mean	63.28	50.32	57.32	56.34	53.71	80.46	77.60	76.19	82.53	78.43
MCC	66.25	51.26	59.60	58.67	55.85	88.22	84.70	83.03	91.04	85.96
J-Statistics	63.06	48.37	56.54	55.63	52.87	84.81	81.31	79.65	87.62	82.57
AUC	83.13	75.63	79.80	79.34	77.92	94.11	92.35	91.51	95.52	92.98

Table 7: Performance measures of rice production for *Rabi* season at Puri district

Measures	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA
Specificity	93.83	93.51	92.78	94.33	95.27	99.54	97.66	97.50	98.99	96.80
Sensitivity	67.25	63.74	56.15	68.38	74.56	98.02	88.64	87.50	95.27	82.55
Precision	67.25	63.74	56.15	68.38	74.56	98.02	88.64	87.50	95.27	82.55
NPV	93.83	93.51	92.78	94.33	95.27	99.54	97.66	97.50	98.99	96.80
FPR	6.17	6.49	7.22	5.67	4.73	0.46	2.34	2.50	1.01	3.20
FNR	32.75	36.26	43.85	31.62	25.44	1.98	11.36	12.50	4.73	17.45
FDR	32.75	36.26	43.85	31.62	25.44	1.98	11.36	12.50	4.73	17.45
F-Score	58.54	55.10	47.66	59.65	65.74	88.95	79.65	78.53	86.22	73.63
G-Mean	56.37	52.81	45.05	57.51	63.78	87.45	78.00	76.86	84.68	71.86
MCC	61.08	57.25	48.93	62.71	69.83	97.56	86.29	85.00	94.25	79.36
J-Statistics	57.55	53.83	45.75	59.16	66.13	93.57	82.36	81.09	90.27	75.51
AUC	80.54	78.63	74.46	81.36	84.92	98.78	93.15	92.50	97.13	89.68

Table 8: Performance measures of rice production for *Kharif* season at Balasore district

Measures	SVM	k-NN	NB	DT	LDA	# SVM	# k-NN	# NB	# DT	# LDA
Specificity	89.83	86.39	87.57	89.35	87.27	98.10	96.99	95.09	96.85	95.38
Sensitivity	68.99	47.60	52.36	60.76	44.26	94.81	91.10	83.60	89.92	83.60
Precision	68.99	47.60	52.36	60.76	44.26	94.81	91.10	83.60	89.92	83.60
NPV	68.99	47.60	52.36	60.76	44.26	94.81	91.10	83.60	89.92	83.60
FPR	89.83	86.39	87.57	89.35	87.27	98.10	96.99	95.09	96.85	95.38
FNR	10.17	13.61	12.43	10.65	12.73	1.90	3.01	4.91	3.15	4.62
FDR	31.01	52.40	47.64	39.24	55.74	5.19	8.90	16.40	10.08	16.40
F-Score	31.01	52.40	47.64	39.24	55.74	5.19	8.90	16.40	10.08	16.40
G-Mean	60.25	39.34	43.96	52.17	36.10	85.77	82.09	74.67	80.92	74.67
MCC	58.13	36.25	41.16	49.76	32.77	84.22	80.48	72.92	79.29	72.92
J-Statistics	58.81	33.99	39.93	50.11	31.53	92.91	88.09	78.69	86.77	78.98
AUC	54.94	30.91	36.63	46.49	28.61	88.59	83.80	74.48	82.48	74.78

Table 9: Performance measures of rice production for *Kharif* season at Cuttack district

Measures	**SVM**	**k-NN**	**NB**	**DT**	**LDA**	**# SVM**	**# k-NN**	**# NB**	**# DT**	**# LDA**
Specificity	92.18	89.96	91.42	91.46	91.22	97.14	96.45	96.18	97.90	96.89
Sensitivity	71.50	57.05	64.88	63.78	60.84	90.32	87.23	85.69	92.57	88.13
Precision	71.50	57.05	64.88	63.78	60.84	90.32	87.23	85.69	92.57	88.13
NPV	71.50	57.05	64.88	63.78	60.84	90.32	87.23	85.69	92.57	88.13
FPR	92.18	89.96	91.42	91.46	91.22	97.14	96.45	96.18	97.90	96.89
FNR	7.82	10.04	8.58	8.54	8.78	2.86	3.55	3.82	2.10	3.11
FDR	28.50	42.95	35.12	36.22	39.16	9.68	12.77	14.31	7.43	11.87
F-Score	28.50	42.95	35.12	36.22	39.16	9.68	12.77	14.31	7.43	11.87
G-Mean	62.73	48.54	56.21	55.14	52.26	81.32	78.25	76.73	83.55	79.15
MCC	60.69	45.97	53.96	52.84	49.85	79.70	76.58	75.03	81.96	77.49
J-Statistics	63.69	47.01	56.30	55.24	52.06	87.47	83.68	81.87	90.48	85.02
AUC	60.32	44.04	53.09	52.06	48.97	83.81	80.05	78.26	86.80	81.38

Table 10: .Performance measures of rice production for *Kharif* season at Puri district

Measures	**SVM**	**k-NN**	**NB**	**DT**	**LDA**	**# SVM**	**# k-NN**	**# NB**	**# DT**	**# LDA**
Specificity	93.83	93.51	92.78	94.33	95.27	99.54	97.66	97.50	98.99	96.80
Sensitivity	78.91	76.87	72.58	79.58	83.29	98.61	92.24	91.49	96.70	88.29
Precision	78.91	76.87	72.58	79.58	83.29	98.61	92.24	91.49	96.70	88.29
NPV	78.91	76.87	72.58	79.58	83.29	98.61	92.24	91.49	96.70	88.29
FPR	93.83	93.51	92.78	94.33	95.27	99.54	97.66	97.50	98.99	96.80
FNR	6.17	6.49	7.22	5.67	4.73	0.46	2.34	2.50	1.01	3.20
FDR	21.09	23.13	27.42	20.42	16.71	1.39	7.76	8.51	3.30	11.71
F-Score	21.09	23.13	27.42	20.42	16.71	1.39	7.76	8.51	3.30	11.71
G-Mean	70.04	68.03	63.79	70.69	74.37	89.53	83.22	82.48	87.64	79.30
MCC	68.18	66.12	61.78	68.86	72.61	88.05	81.63	80.88	86.13	77.65
J- Statistics	72.74	70.38	65.36	73.91	78.56	98.15	89.89	88.99	95.69	85.09
AUC	70.03	67.70	62.74	71.19	75.80	95.30	87.06	86.16	92.84	82.28

classifier gives better performance. In other performance cases also DT gives better. So it is seen that in case of Kharif season also DT gives better as in case of Rabi season of Cuttack district.

In **Table 10** it is seen that, the performance of specificity improves from 96.80% to 99.54%, 88.29% to 98.61% in case of sensitivity, precision and NPV. Here also SVM main classifier gives better performance than others. Also in case of all other performance measures, SVM gives better result than other main classifiers.

By summarizing the result part we can get that the main classifier of the ensemble method gives better result than the individual classifier. From all the main classifiers, the SVM gives better in case of Balasore and Puri district but DT gives better result in case of Cuttack district. But DT result is not more enough than the SVM. So we can conclude that, when we are considering SVM as main classifier then getting better result. So overally it concludes that the ensemble method gives better performance than the individual classifiers.

7 Discussions

This work aimed at development of a computational model for prediction of rice yield by considering the effect of climatic variability for the coastal state of India i.e. Odisha. The districts such as Balasore, Cuttack and Puri were considered for Rabi and Kharif seasons. For experimentation we have used five classifiers such as SVM, k-NN, NB, DT and LDA. The following points summarize this work;

- The datasets were first constructed and class labels are identified.
- The window size for training is chosen for Rabi ($w = 10$) and Kharif ($w = 12$) season datasets experimentally.
- A multi-classifier based ensemble model has been proposed where, four classifiers are chosen as base classifiers and the output of those base classifiers ($\hat{S}$) are passed though the conversion function f as to the main classifier.
- The main classifier containing the dataset augmented with the output of base classifiers is used for the prediction.
- The result obtained after processing augmented dataset by main classifier is compared expected output.
- Final prediction is made by using majority voting on the class label predicted by each classifier as main classifier.
- The effectiveness of proposed model has been verified by measuring the average classification accuracy for all the individual classifiers, main classifiers and the final result obtained after majority voting.

- It can be observed from **Table 3** and **Table 4** that, the average classification accuracy obtained after majority voting is above 96% for both *Rabi* and *Kharif* season datasets, because in this model it considers the best classifiers predicted output for finding the final predicted output.
- It is also evident that, the improvements of performance measures approaches towards 95.09% to 98.10% and 95.38% to 98.10% for specificity, 88.48% to 96.25% and 83.60% to 94.81% for both sensitivity and precision and 91.78% to 97.17% and 74.48% to 88.59% for AUC for Rabi and Kharif seasons dataset of Balasore district which is observed in the **Table 5** and **Table 8**.
- The improvements of performance measures are 96.45% to 97.14% and 96.18% to 97.90% for specificity, 86.85% to 93.14% and 87.23% to 92.57% for both sensitivity and precision and 91.51% to 95.52% and 78.26% to 86.80% for AUC for *Rabi* and *Kharif* seasons dataset of Cuttack district described in the **Table 6** and **Table 9**.
- Similarly, the improvements of performance measures are 96.80% to 99.54% and 96.80% to 99.54% for specificity, 82.55% to 88.64% and 88.29% to 98.61% for both sensitivity and precision and 89.68% to 98.78% and 82.28% to 95.30% for AUC for *Rabi* and *Kharif* seasons dataset of Puri district which can be observed in the **Table 7** and **Table 10**.

8 Conclusion

Due to variation in temperature, humidity, precipitation and other metrological variable in a particular area for a period of at least 25 years the expected crop yields are not produced in India. Odisha's climate has also under gone appreciable changes due to various factors. The deviation in day temperature and annual rain fall is mostly restricted to 4 months in a year and number of rain days decreased from 120 to 90 days besides being erratic. In addition, the mean temperature is increasing and minimum temperature has increased about 25 %. Such climate change related adversity is affecting adversely productivity and production of food grains. The production of rice mainly depends on climatic parameters such as rainfall, temperature, humidity, wind speed etc. If the farmers will be able to know the timely variation in climatic conditions they can get the timely advice to increase the production. Therefore, in this work we have proposed machine learning based multi-classifier approach of ensemble learning mechanism using majority voting approach to predict the rice yield based on thirty years rice production as well as climate original datasets. Our model shows above 96% classification accuracy and also the performance of the proposed model has been compared with individual classifiers and shows that the main classifier gives better result than the individual classier. Additionally, the classifier's accuracy validity and statistical test are conducted to establish the performance of the model. This model can give prediction value of the rice production, but can't explain which parameter affect mostly for the production. This limitation can be extended by the researcher. This ensemble based prediction model can also be extended for prediction of different crop yield.

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] Hansen J. W., and Sivakumar M. V. (2006). Advances in applying climate prediction to agriculture. *Climate Research*, 33(1), 1-2.

[2] Rattan R.K., (2014). "Soil process and climate change", *delivered on 25th November at the 79th Annual convention of Indian Society of Soil Science*, PJTSAU, Hyderabad.

[3] Venkateswarlu B. (2010). "Adaptation and mitigation strategies in rain fed agriculture", *Journal of the Indian Society of Soil Science* 58, S27-S35.

[4] Saseendran A.S.K., Singh K.K., Rathore L.S., Singh S.V. and Sinha S.K., (2000) "Effect of climate change in rice production in the tropical humid climate of kerala, india", Climate Change, 44, pp. 495-514.

[5] Lal M., Singh K.K., Rathore L.S., Srinivasan G. and Saseendran S.A.,(1998) "Vulnerability of rice and wheat yield in NW India to future changes in climate", Agricultural and Forest meteorology, 89, pp.101-104.

[6] Reddy V.R. and Pachepsky Ya.A., (2000) "Predicting crop yields under climate change conditions from Monthly GCM weather projections", Environmental Modelling and Software, 15, pp.79-86.

[7] Narayanan B, Govindarajan M. (2016) "Crop Production-Ensemble Machine Learning Model for Prediction", International Journal of Computer Science and Software Engineering (IJCSSE), 5(7), pp.148-153.

[8] Narayanan B., Govindarajan M. (2016) "Rainfall Prediction based on Ensemble Model", International Journal of Innovative Research in Science, Engineering and Technology, 5(5).

[9] Sadegh B.I., Mohammad B. (2013) "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background", Int. Journal of Engineering Research and Applications, 3(5), pp.605-610.

[10] Devroye L. , (1981) "On the equality of Cover and Hart in nearest neighbor discrimination", IEEE Trans. Pattern Anal. Mach. Lntell, 3,pp. 75- 78.

[11] Devroye L., Gyorfi L., Krzyzak A. and Lugosi G., (1994) "On the strong universal consistency of nearest neighbor regression function estimates", Ann. Statist, 22, pp. 1371– 1385.

[12] Devroye L. and Wagner T.J., (1982) "Nearest neighbor methods in discrimination, In Classification, Pattern Recognition and Reduction of Dimensionality", Handbook of Statistics, 2: pp. 193–197.

[13] Cover T.M. (1968) "Rates of convergence for nearest neighbour procedures", In Proceedings of the Hawaii International Conference on System Sciences, Univ. Hawaii Press, Honolulu, pp. 413–415.

[14] Devroye L.(1981) "On the asymptotic probability of error in nonparametric discrimination", Ann. Statist, 9, pp. 1320–1327.

[15] YiqunGu Y., James W., McNicol M. (1994) " An Application of Belief Networks to Future Crop Production", IEEE conference on Artificial Intelligence for Applications, San Antonia, TX , pp.305-309.

[16] Uno Y. (2005) "Artificial Neural Networks to Predict Corn Yield from Compact Airborne Spectographic Imager Data", Computers and Electronics in Agriculture, 47(2), pp.149-161.

[17] Veenadhari S., Mishra B., Singh C.D. (2011) "Soybean Productivity Modelling using Decision Tree Algorithms", International Journal of Computer Applications, 27(7), pp. 975-8887.

[18] Veenadhari S., Misra B., Singh C. D. (2014) "Machine learning approach for forecasting crop yield based on climatic parameters", IEEE International Conference on Computer Communication and Informatics , Coimbatore, pp. 1-16.

[19] Alexander B., Klaus K. and Itzerott S. (2006) "Comparing Classifiers For Crop Identification Based On Multitemporal Landsat Tm/Etm Data", Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover, Centre for Remote Sensing of Land Surfaces, Bonn, pp. 28-30.

[20] Minggang D. W. (2011) "Linear Discriminant Analysis and Its Application in Plant Classification", ICIC '11 Proceedings of the 2011 Fourth International Conference on Information and Computing, Pages pp. 548-551.

[21] Renfang L. (2016) "Using Penalized Linear Discriminant Analysis and Normalized Difference Indices Derived from Landsat 8 Images to Classify Fruit tree Crops in the Aconcagua Valley, Chile", A thesis presented to the University of Waterloo in fulfilment of the thesis requirement for the degree of Master of Science in Geography, Waterloo, Ontario, Canada.

[22] Chen Yud-Ren, Chao, Kuang L. and Kim S. M. (2002) "Machine vision technology for agricultural applications", Computers and Electronics in Agriculture, 36, pp.173-191.

[23] Olson M. Jennifer; Alagarswamy G., Andresen J. A., Campbell D. J., Davis A. Y., Ge J. et al. (2008) "Integrating diverse methods to understand climate–land interactions in East Africa", Geoforum, 39, pp. 898–911.

[24] Everingham Y.L.; Smyth C.W. and Inman-Bamber N.G. (2009) "Ensemble data mining approaches to forecast regional sugarcane crop production", Agricultural and Forest Meteorology. 149, pp.689-696.

[25] Wang, N., Zhang N., and Wang M. (2006) "Wireless sensors in agriculture and food industry—Recent development and future perspective", Computers and Electronics in Agriculture, 50, pp.1–14.

[26] Huang Y., Lan Y., Thomson J.S., Fang A., Wesley C. and Lacey E. R. (2010) "Development of soft computing and applications in agricultural and biological engineering.",Computers and Electronics in Agriculture, 71,pp. 107-127.

[27] Wang L. (2010) "Support Vector Machines: Theory and Applications", Springer-Verlag, Berlin Heidelberg.

[28] Han J., Kamber M. (2006) "Data Mining Concepts and Techniques", Elsevier Science and Technology, Amsterdam.

[29] Vapnik V., Lerner A. (1963) "Pattern recognition using generalized portrait method", Automat Remote Contr, 24,pp.774-780.

[30] Smola A., Schlkopf B.(2004) "A tutorial on support vector regression", StatComput 14(3), pp. 199-222.

[31] Vapnik V., Golowich S., Smola A.(1997) "Support vector method for function approximation, regression estimation, and signal processing", MIT Press, Cambridge, MA, USA, pp. 281-287.

[32] Kumar R., Singh M.P., Kumar P. and Singh J.P. (2015) " Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique" , 2015 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R and D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015, pp.138-145.

[33] Fix E., Hodges J.L. (1951) "Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties", USAF school of Aviation Medicine, Randolph Field Texas.

[34] Maria Rossana C. de Leon, Eugene Rex L. Jalao, (2013) "A Prediction Model Framework for Crop yield Prediction", Asia Pacific Industrial Engineering and Management System Conference.

[35] Larose, D.T., (2005) "Discovering Knowledge in Data: An Introduction to Data Mining", Wiley, Chichester.

[36] Zahoor J., Abrar M., Bashir S., and Mirza A.M. (2008) "Seasonal to Inter-annual Climate Prediction Using Data Mining KNN Technique", IMTIC 2008, CCIS 20, pp. 40–51, 2008. © Springer-Verlag Berlin Heidelberg.

[37] Sharma M. and Sharma S.K. (2013) "Generalized K-Nearest Neighbour Algorithm- A Predicting Tool", International Journal of Advanced Research in Computer Science and Software Engineering, 3(11), November.

[38] Mishra S., Mishra D., Mallick P.K., Santra G.H. and Kumar S. (2021), A Novel Borda Count based Feature Ranking and Feature Fusion Strategy to Attain Effective Climatic Features for Rice Yield Prediction, Informatica, 45(1), pp. 13-31.

[39] Korada N.K., SagarPavan N.K. and Deekshitulu Y.V.N.H. (2012) "Implementation of Naive Bayesian Classifier and Ada-Boost Algorithm Using Maize Expert System", International Journal of Information Sciences and Techniques (IJIST), 2(3).

[40] Vijay S., Solanki S. (2014) "Data Mining Techniques Using WEKA classification for Sickle Cell Disease", (IJCSIT) International Journal of Computer Science and Information Technologies, 5(1),pp. 1-26.

[41] Sebastian R. (2014) "Linear Discriminant Analysis-Bit by Bit", Articles, Aug 3.

[42] Alexander B. Statistical Geocomputing, (2013) "Statistical and Machine-Learning Classification Methods", University of Waterloo, Canada, GEOSTAT.

[43] Robi P. (2009) "Ensemble learning", Scholarpedia, the peer reviewed open access encyclopedia, 4(1), pp. 2776.

[44] Sarwesh S., Sadhna K. M. (2013) " A Review of Ensemble Technique for Improving Majority Voting for Classifier", International Journal of Advanced Research in Computer Science and Software Engineering, 3(1). pp-177-180.

[45] Xueyi W. (2012) "A New Model for Measuring the Accuracies of majority voting ensembles", IEEE World Congress on Computational Intelligence.

[46] Orissa Agricultural Statistics Year Book, 1983-2013. Published by Directorate of Agriculture and Food Production, Govt. of Odisha, Bhubaneswar.