

## A Hybrid Approach of TLBO and EBPNN for Crop Yield Prediction Using Spatial Feature Vectors

Preeti Tiwari<sup>1,\*</sup> and Piyush Shukla<sup>1</sup>

**Abstract:** The prediction of crop yield is one of the important factor and also challenging, to predict the future crop yield based on various criteria's. Many advanced technologies are incorporated in the agricultural processes, which enhances the crop yield production efficiency. The process of predicting the crop yield can be done by taking agriculture data, which helps to analyze and make important decisions before and during cultivation. This paper focuses on the prediction of crop yield, where two models of machine learning are developed for this work. One is Modified Convolutional Neural Network (MCNN), and the other model is TLBO (Teacher Learning Based Optimization)-a Genetic algorithm which reduces the input size of data. In this work, some spatial information used for analysis is the Normalized Difference Vegetation Index, Standard Precipitation Index and Vegetation Condition Index. TLBO finds some best feature value set in the data that represents the specific yield of the crop. So, these selected feature valued set is passed in the Error Back Propagation Neural Network for learning. Here, the training was done in such a way that all set of features were utilized in pair with their yield value as output. For increasing the reliability of the work whole experiment was done on a real dataset from Madhya Pradesh region of country India. The result shows that the proposed model has overcome various evaluation parameters on different scales as compared to previous approaches adopted by researchers.

**Keywords:** Crop yield prediction, data mining, machinelearning, vegetation index, TLBO.

### 1 Introduction

The number of inhabitants on the planet is expanding day by day, so for the better nourishment and administration, the need for a far-reaching, dependable and quick data resources are required. The utilization of remote detecting has ended up being exceptionally critical in areas of agriculture for observing the development of crops and irrigation system. Remote sensing information can be utilized for various applications, for example, crop stock, crop generation and dry season prediction, and irrigation land checking.

The objective of the information mining process is to remove learning from a current information index and change it into a human reasonable development for higher utilization [Pritam and Kasabov (2016)]. Information mining is an incredible request in other handy databases, including spatial databases, worldly databases, object arranged databases, multimedia databases, so on. Spatial information mining can be utilized to anticipate crop

---

<sup>1</sup> UIT, RGPV, Bhopal, 462033, India.

<sup>2</sup> UIT, RGPV, Bhopal, 462033, India.

\* Corresponding Author: Preeti Tiwari. Email: preetitw77@gmail.com.

yield expectation other than genuine applications. Spatial information mining is the way toward removing intriguing learning from spatial databases. This learning can be utilized for understanding spatial and non-spatial information with their connections. Information found from spatial information can be of different structures, similar to the trademark and segregate principles, extraction and depiction of unmistakable structures or groups, spatial affiliations [Gandhi, Armstrong, Petkar et al. (2016)].

The test has been to extricate information from this crude information, has prompt new strategies and methods, for example, information mining that can connect the learning of the information to the crop yield estimation [Geeta (2015)].

Crop yield forecast has been a subject of enthusiasm for researchers, specialists, and field associations. As characterized by the Food and Agriculture of the United Nations, crop prediction is the craft of anticipating crop yields and creation before the harvest happens, commonly a few months ahead of time [Ramesh and Vardhan (2015)]. The crop yield forecast is a noteworthy segment of national nourishment security assessment and sustenance approach making. Crop development and yield information are basic for directing rural development framework, and farming task with administration [Medar and Rajpurohit (2014)]. The crop generation could be influenced straightforwardly and by implication under environmental change [Meroni, Fasbender, Balaghi et al. (2016)]. Measurable models and harvest models are two primary devices for contemplating the impacts of environmental change on crop yields [Garg and Bindu (2017)]. Larger part of research works in the farming spotlight on natural instruments to distinguish crop development and enhance its yield. The result of harvest yield fundamentally relies upon parameters, for example, the assortment of the crop, seed write and natural parameters, for example, daylight, soil, water, precipitation and dampness [Kaur, Gulati and Kundra (2014)] Harvest yield estimation is important, especially in nations like India, which relies upon farming as their principle wellspring of the economy. Harvest yield estimation causes choices creators to take choice about the abundance or shortage of generation conditions and permit suitable import and fare choices [Dahikar and Rode (2014)]. Crop recognizable proof and forecast of yield are the principle worry of remote detecting application in agribusiness [Raorane and Kulkarni (2013)].

Spatiotemporal information mining can be utilized to foresee crop yield forecast other than other certifiable applications. This paper tosses light into crop yield forecast utilizing socio temporal information mining and extends it to utilize Map Reduce system with a specific end goal to enhance the prediction precision value. Since the information for spatial information mining is exceptionally huge, information mining can help in extricating thorough business knowledge as data represents the information with attributes of volume, speed and variety.

## **2 Related work**

Ramesh et al. [Ramesh and Vardhan (2015)] analyzes the consequences of different straight regression and Density-based group system. The information was thought about in the particular area i.e., East Godavari region of Andhra Pradesh in India. Numerous Linear Regression is connected on existing information yet the outcomes acquired are analyzed and inspected utilizing Density-based grouping procedure.

Raorane et al. [Raorane and Kulkarni (2013)], recommends that few changes in the climate can be analyzed by Support Vector Machine (SVM is fit for arranging information tests in two disjoint bunches) and furthermore K-implies technique was utilized to forward the contamination in the environment. Information mining strategies are utilized to screen the wine fermentation.

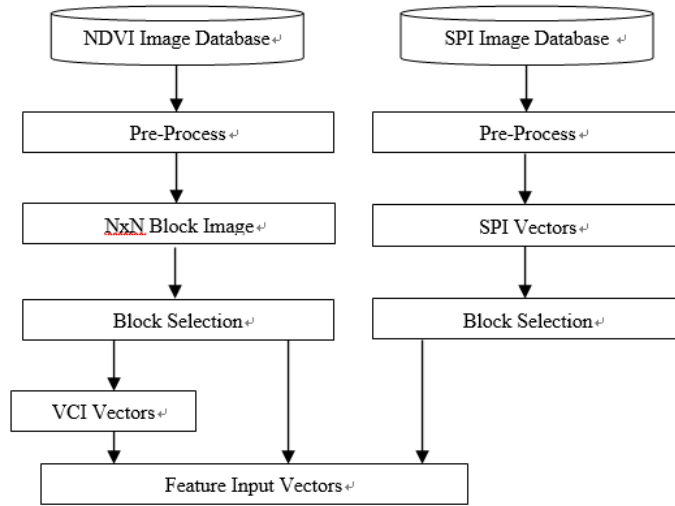
Tits et al. [Tits, Somers and Coppin (2012)] proposed checking framework for crop creation through grouping approach. The sub-pixel cover part and the unconstrained spatial mark of the plant include from the consolidated hyperspatial flag was extricated at the same time by building up different end part phantom mix examination. The age of lookup tables (LUT) was accomplished through a radiative exchange show for both soil and plant features. In the wake of dividing the features, the grouping strategy was performed for upgrading the effectiveness of the use of LUT in this proposed display. At that point, the Bayesian choice calculation was displayed for choosing the most positive bunches. Be that as it may, the impacts of layout in the mix and non-straight mixed impacts in orchard frameworks were not considered.

Iordache et al. [Iordache, Tits, Plaza et al. (2014)] created a dynamic unmixing model for plant generation checking framework. The watched spectra, for example, vegetation and soil were thought to be straight mixes of spectra from accessible otherworldly libraries through direct mixing model (LMM). The unmixed issues were handled by different end part otherworldly mix examination (MEMSA) and inadequate unmixing by means of variable part and increased Lagrangian (SUnSAL). For powerful library lessening, change of Hyper-spatial unmixing by means of numerous flag characterization and synergistic scanty regression (MUSIC-CSR) was created. This procedure was utilized for pruning the word reference and seeing the high calibre of the vegetation spectra on the ground by methods for pruned lexicon as a contribution to accessible unmixing strategies. Notwithstanding, this technique was troublesome for obtaining solid estimation of ground partial plenitudes.

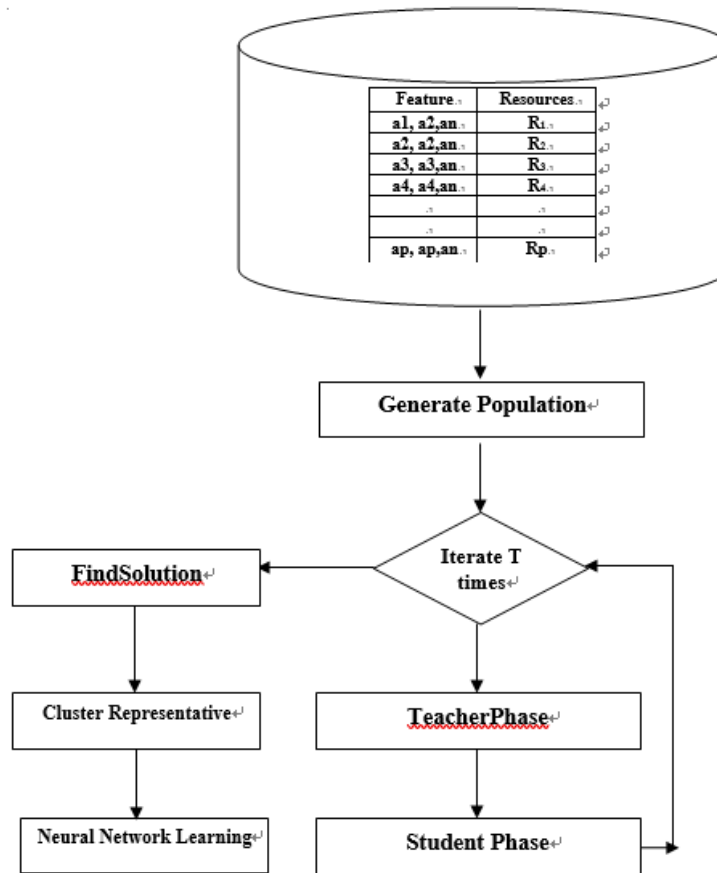
Sung et al. [Sung, Chung and Chang et al. (2014)] explored insect state calculation with focus information total for horticultural checking framework. The plant development was observed through natural parameters gathered by means of Zigbee based climate stations. The whole sensors were consolidated into climate stations and essentially single observing hub was utilized for information total. A vitality effective focus information accumulation technique was exhibited in which subterranean insect province improvement calculation was connected to the creation of level slope field. The remote online human-machine interface was likewise created for checking the plant generation. Notwithstanding, the normal postponement of this system was high.

### **3 Proposed methodology**

In order to make a general model which work on various available data Indices a Teacher Learning Based Optimization (TLBO), a genetic algorithm was used where data is classified without finding all possible combination. In this work, EBPNN is used as the learning model where input data was filtered/refined first by TLBO algorithm. Fig. 1 and Fig. 2 explain the whole working, where each block diagram was explained by the various section.



**Figure 1:** Block diagram of proposed Feature Collection



**Figure 2:** Block diagram of proposed work

### 3.1 Vegetation condition index (VCI)

As suggested in Sung et al. [Sung, Chung, Chang et al. (2014)], which represent a correlation between the current NDVI values with the minimum NDVI value from a huge set of period. It was calculated as shown in Eq. (2).

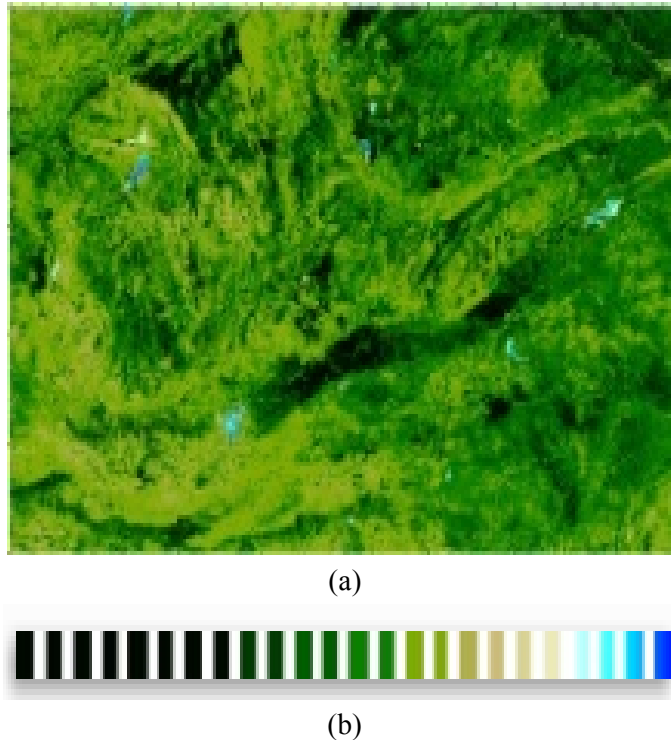
$$V_j = \left( \frac{N_j - N_{\min}}{N_{\max} - N_{\min}} \right) \times 100 \quad (1)$$

where  $V_j$  stands for the  $j^{\text{th}}$  position VCI value in the vector. Similarly  $N_j$  stands for the  $j^{\text{th}}$  position NDVI value in the vector. While  $N_{\min}$  and  $N_{\max}$  are minimum and maximum values of NDVI vector.

### 3.2 Pre-processing

Input image data need to be preprocessed first for transforming data as per the proposed model. It involves activities like cleaning of data and converts in required environment format. This can be understood as the let image consist of four pixels having the dimension of  $2 \times 2$  then for this image, a matrix is of same dimension  $2 \times 2$  and its four-cell contain value as per the pixel colour and representing format. In this step, image is resized in fix dimension. As different image has different dimension. So conversion of each is done in this step.

### 3.3 Index vector



**Figure 3:** Input data for getting NDVI Vector, (a) NDVI Image, (b) NDVI scale

In order to get index value from the image, each image has its own colour scale where pixel value corresponding to a specific colour is replaced by its index value. This can be understood as let Fig. 3(a) has an NDVI index than all the dark green portion in the image is considered as the extremely wet region, so if this dark green is represented by pixel value 65 then all the pixel value in the image of NDVI where 65 value is present is considered as extremely wet region and NDVI value for the cell is +1.0. In similar fashion other regions are also detected in the image. While SPI image index value is obtained from its separate index, as the image colour region of this image is different.

### **3.3 Feature selection**

In this work, three indexes are used as the training parameters, first was NDVI, the second was SPI and third was VCI. Use of three feature vectors for training increase the accuracy of the resultant neural network. As this help in covering various other geographical parameters of the location. So a set of values for any geographical location at a particular period of time this feature vector has three value set. This can be understood as if the geographical location  $(x_i, y_i)$ , then its corresponding feature vector for time instance  $t_i$  is  $[N_i, V_i, S_i]$ . While for same location but at time  $t_j$ , it might be  $[N_j, V_j, S_j]$ .

So for each year data input feature vector is a three-dimension vector where the first and second dimension provide geographical position while third provide index vector for the same.

### **3.4 TLBO (Teacher learning based optimization)**

In this model, TLBO (Teachers Learning Based Optimization Algorithm) was used for classifying the feature value of NDVI, VCI, SPI. Here this reduces the size of the dataset as during learning presence of similar data with multiple value increase the confusion and resultant value will be deflected. In this work TLBO, a genetic algorithm is used because this takes two-phase learning. The main motive of this model is to reduce the dataset size and increase the leaning accuracy of the neural network. Here iteration of teacher and student phase was done while two similar solutions were not obtained.

### **3.5 Generate population**

In this step cluster center set was generate, which represent whole input block of the image. So each cluster center set acts as the chromosome while the collection of all set is termed as population. This can be assumed as  $C = [c_1, c_2, \dots, c_m]$  as the cluster center set. While  $P = [C_{c1}, C_{c2}, \dots, C_{cn}]$  where  $m$  is the number of cluster and  $n$  is the number of chromosomes. In this work, each block is divided into  $m$  cluster, where each cluster center has NDVI, VCI and SPI value of the selected pixel position.

### **3.6 Fitness function**

In order to obtain good chromosome from the bunch of available set fitness value of each probable cluster center set is passed in this function. So the fitness value was returned. Now applying Euclidian distance formula in Eq. (2) is used for finding the fitness value. This can be understood as let cluster set  $C_c = [C_1, C_2]$  fitness value need to calculate. Than distance of cluster center pixel values from each non-cluster center pixel of the

block was evaluate. Now each non-cluster center pixel was assigned to the minimum distance cluster center. While the sum of all minimum distance of each non-cluster center was considered as fitness value shown in Eq. (3).

$$D_{m,s} = \sum \sqrt{(Cc_i - B_s)^2} \tag{2}$$

$$F = \sum \min(D_{1,s}, D_{2,s}, \dots \dots \dots D_{m,s}) \tag{3}$$

where m is the number of cluster center and s is the number of pixel in the block of image.

**3.7 Teacher phase**

This phase was used for the crossover of the chromosomes by the single best solution from the population. Here best solution act as a teacher and its selection is based on the minimum fitness value. In order to do crossover operation, random position cluster center value is copied from the teacher chromosome and it was replaced to the non-teacher chromosome. This improves the population quality. This can be understood as: let the best solution is Ccb, then the crossover operation done through Eq. (4).

$$Cc [m, r] \leftarrow Cc_b[b, r] ; \text{where } r=[1 \text{ to } n] \tag{4}$$

**3.8 Student phase**

In this phase, some random group of the chromosome were made automatically and then each group was used for the crossover of the chromosomes by the single best solution in that group. Here best solution act as a teacher among other chromosomes and its selection is based on the minimum fitness value. In order to do crossover operation, random position cluster center value was copied from the teacher chromosome and it was replaced to the non-teacher chromosome using Eq. (4). Here each new chromosome was cross-verified that either its fitness value gets improved from then previous one, if fitness improves, then the new chromosome is included in the population and older one get removed and vice versa if fitness value not improves.

**3.9 Training vector**

Here the output of the genetic algorithm reduces the feature vector as each block was represented by cluster center value of NDVI, SPI and VCI. So each cluster center acts as input values and crop yield of that area in that year is considered as the desired output. So the input and output parameters are combined to generate training data for the Error Back Propagation Neural Network.

**3.10 Training of error back propagation neural network (EBPNN)**

- Let us assume a three layer neural network.
- Now consider i as the input layer of the network. While j is consider as the hidden layer of the network.
- Finally k is considered as the resultant output layer of the network.
- If  $w_{ij}$  represents a weight of the between nodes of different consecutive layers.
- So the output of the neural network is depending on the below equation sigmoidal function shown in Eq. (5):

$$Y_j = \frac{1}{1+e^{-x_j}} \quad (5)$$

where,  $X_j = \sum x_i \cdot w_{ij} - \theta_j$ ,  $1 \leq i \leq n$ ;  $n$  is the number of inputs to node  $j$ , and  $\theta_j$  is threshold for node  $j$ .

In order to understand above steps let us consider an example where  $W_{ij}$  have some weight values.

$$W_{ij} = \begin{vmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{vmatrix}$$

Let input vector of three value which include values from NDI, VCI, SPI images are pass from the first layer of neural network. These values get multiply by above weight values.

Now, this act as input  $H1_{input}$  to next layer of hidden neurons. In this, some biasing is also possible which was neglected in this example. So weight values of the neuron for next level is assumed as shown in the below matrix.

$$W_{jk} = \begin{vmatrix} W'_{11} & W'_{12} & W'_{13} \\ W'_{21} & W'_{22} & W'_{23} \\ W'_{31} & W'_{32} & W'_{33} \end{vmatrix}$$

where each value obtained from the previous weight matrix multiplication is passed through the sigmoidal function as in Eq. (5) Therefore, a small variation in the output value was done by this function.

Let the expected output vector be  $y=[0 \ 1 \ 0]$  and obtained values are  $O=[0.6536, 0.7197, 0.6761]$  so difference between the expected with obtained is consider as the error. This error needs to be corrected by adjusting the weight values of each layer. Here the forward movement of the neural network is over and error backpropagation (EBPNN) starts by Eq. (6).

$$\frac{\partial E_i}{\partial O_i} = \frac{\partial(-1 * ((y_i * \log(O_i) + (1 - y_i) * \log(1 - O_i)))}{\partial O_i}$$

$$\frac{\partial E_i}{\partial O_i} = (-1 * ((y_i * \log(O_i) + (1 - y_i) * \log(1 - O_i))) \quad (6)$$

In similar fashion other values can be calculated to find other set of derivatives for sigmoid of Eq. (7).

Here as per output derivative value may vary.

$$\frac{\partial O_i}{\partial H_i} = \frac{\partial\left(\frac{1}{1+e^{-x}}\right)}{\partial x} = \left(\frac{1}{(1+e^{-x})} \times (1 - (1/(1+e^{-x})))\right) \quad (7)$$

For each input to neuron let us calculate the derivative with respect to each weight. Now let us look at the final derivative:

$$\sum_{i=1:n} \frac{\partial H_i}{\partial W_{i(j,k)}} = \frac{\partial(h_{i(ouput)} * W_{i(j,k)})}{\partial W_{i(j,k)}} \quad (8)$$



Now by using chain rule final derivatives were calculated for the same. Here multiplication of output obtained from Eqs. (6)-(8) was done in following way:

$$\frac{\partial E_i}{\partial W_i} = \frac{\partial E_i}{\partial O_i} * \frac{\partial O_i}{\partial H_i} * \frac{\partial H_i}{\partial W_i} \quad (9)$$

So the overall  $\partial W_i$  can be obtained by getting the value of weight from the above equation, here all set of weight which needs to be updated are change by below matrix values.

$$\partial W_i = \begin{bmatrix} \frac{\partial E_1}{\partial W_{1,1}} & \frac{\partial E_2}{\partial W_{1,2}} & \frac{\partial E_3}{\partial W_{1,3}} \\ \frac{\partial E_1}{\partial W_{2,1}} & \frac{\partial E_2}{\partial W_{2,2}} & \frac{\partial E_3}{\partial W_{2,3}} \\ \frac{\partial E_1}{\partial W_{3,1}} & \frac{\partial E_2}{\partial W_{3,2}} & \frac{\partial E_3}{\partial W_{3,3}} \end{bmatrix}$$

- So the error corresponds to the input data was estimated by differencing desired output obtained from the output layer.

$$e_k(n) = d_k(n) - y_k(n)$$

- The ebpnn weight updation was done by the above matrix of  $\partial W_i$

$$W_{ij} = W_{ij} + \Delta W_{ij}$$

So at the end of the above iteration steps are over when error obtained from the output layer get nearer to zero or some constant such as 0.001.

### **3.11 Testing of EBPNN**

Testing of Trained neural network obtained from the above steps is processed by passing testing input images of NDVI, SPI. These images are preprocessed and generate feature values of as done in training part of the neural network. Finally, feature vectors were passed in the EBPNN which gives crop yield predicted value. This crop yield predicted values are compared on various parameters to find the fitness of the proposed work.

## **4 Experiment and results**

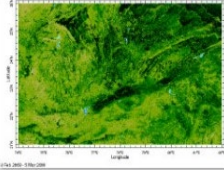
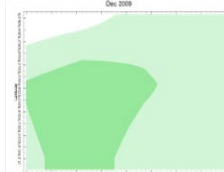
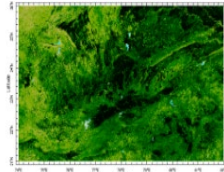
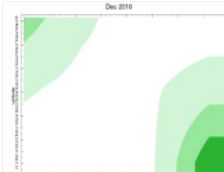
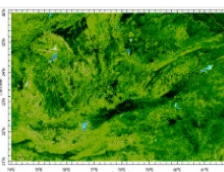
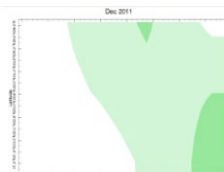
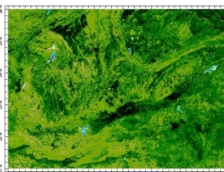
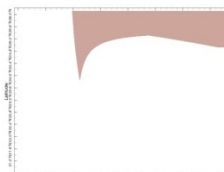
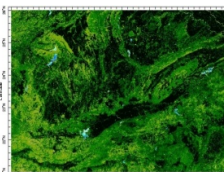



In order to conduct an experiment and to measure evaluation results, MATLAB 2012a version software is used. This section of the paper shows the experimental setup and results. The tests were performed on a 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional.

### **Dataset**

The experiment is done on a real dataset having the SPI from one of the world wide web source: <https://iridl.ldeo.columbia.edu/maproom/Global/Precipitation/SPI.html>. While NDVI value obtained from <http://iridl.ldeo.columbia.edu/SOURCES>. Each source image is extracted on the longitude parameters 21N-26N 60, and Latitude parameters 74E-82E, where the selected region is one of state Madhya Pradesh in India. While the average of different crop yields is obtained from <http://www.mospi.gov.in/statistical-year-book->

[india/2016/177](http://www.mospi.gov.in). Ground truth values are also considered from the same site <http://www.mospi.gov.in> for testing of trained models.

**Table 1:** Dataset NDVI and SPI values

Year	NDVI Image	SPI Image
2009		
2010		
2011		
2012		
2013		
2014		

## 5 Results

Results found are compared with the previous work [Spiking Neural Network (SNN) in Pritam et al. [Pritam and Kasabov (2016)]], which is termed as previous work in this paper.

**Table 2:** RMSE value comparison of previous work with proposed work different sets

Year	RMSE		
	SNN	EBPNN	GA+EBPNN
2012	2509.4	943.7	4.0909
2013	1401.1	1188.1	419.2222
2014	1615.3	1631.8	442.1667

It has been observed by Tab. 2 that proposed work GA+EBPNN of crop yield prediction using multiple feature work well as compare to the previous method adopt in Pritam et al. [Pritam and Kasabov (2016)] and EBPNN. Here RMSE value of proposed work is low as compared to the spiking neural network model. In this work, the use of TLBO algorithm for reducing input feature data increase the value of accuracy as well.

**Table 3:** Relative error value comparison of previous work with proposed work different sets

Year	Relative Error		
	SNN	EBPNN	GA+EBPNN
2012	19.1075	7.1858	0.0311
2013	10.8299	9.1835	3.2405
2014	11.5063	11.3895	3.1178

It has been observed by Tab. 3 that relative error of the proposed work of crop yield prediction work is less as compared to the previous method adopt in Pritam et al. [Pritam and Kasabov (2016)]. One more important factor is observed that by the change in climate production of crops also rises. Above table shows that the combination of two soft computing techniques from the genetic and neural network has reduced Relative error value.

**Table 4:** Yield in Tones value comparison of previous work with proposed work different sets

Year	Yield in Tones Values			
	Ground Truth	SNN	EBPNN	GA+EBPNN
2012	13133	10624	12196	13130
2013	12937	11536	11786	13355
2014	14182	12550	12567	13742

It has been observed by Tab. 4 that the relative error of the proposed work of crop yield prediction value is much nearer to the ground truth value as compared to the previous method adopt in Pritam et al. [Pritam and Kasabov (2016)]. As dimension reduction done by TLBO, a good set of cluster center were selected and reduce the confusion of neural network while training.

**Table 5:** Training time value comparison of previous work with proposed work different sets

Iterations	Training Time in seconds		
	SNN	EBPNN	GA+EBPNN
10	29.36	15.23	14.905
20	42.58	26.98	18.2541
40	74.46	48.23	25.2667

It has been observed by Tab. 5 that training time required for error back propagation neural network is comparatively low because of less number of neuron in input layer as compare to spiking neural network used in previous work [Pritam and Kasabov (2016)]. Here the input layer in the EBPNN depends on the number of features used, while in case of spiking neural network number of neurons in input layer depend on training block size. While in case of GA+EBPNN it was assumed that finding cluster center is a kind of pre-processing step before the input in the neural network training. So, based on that execution time of proposed work was found to be less as compared to other existing methods.

## 6 Conclusion

Crop yield prediction through various machine learning techniques can be helpful to agricultural departments to have strategies for improving agriculture. Nowadays, various techniques are coming into existence for the same cause of future prediction. Spatiotemporal data mining is one such solution that can be employed to achieve crop yield prediction. Considering this fact, this paper has focused on one of the issues of the crop yield geospatial based prediction by using different indices such as NDVI, SPI, VCI. Many researchers have already done a lot of work based on neural network classification. In this model, TLBO (Teachers Learning Based Optimization Algorithm) was used for classifying the feature value of NDVI, VCI, SPI. Here the reduce in the size of the dataset is done during the learning phase- as the presence of similar data with multiple values increases the confusion and the resultant value gets deflected. In this work, the TLBO, genetic algorithm is used because this takes two-phase learning. The main motive of this model is to reduce the dataset size and increase the leaning accuracy of the neural network. As during learning in ANN various values to the same output increases the confusion, so the overall output gets disturbed. Here, the result shows that proposed work has improved the prediction accuracy by reducing the RMSE, relative error value. Also through proper training and rich input vector, the resultant neural network is less time consuming model by 57.47% as compared to SNN. It was obtained that the proposed work has reduced the relative error by 80.93% as compared to SNN, while RMSE also gets reduced by 80.85%. Hence the overall accuracy of the proposed work was also improved.

## References

**Ankalaki, S.; Chandra, N.; Majumdar, J.** (2016): Applying data mining approach and regression model to forecast annual yield of major crops in different district of karnataka. *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, pp. 25-29.

**Bose, P.; Kasabov, N. K.; Bruzzone, L.; Hartono, R. N.** (2016): Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 6563-6573.

**Dahikar, S.; Rode, S.** (2014): Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronic Instrumentation and Control Engineering*, vol. 2, no. 1, pp. 683-686.

**Gandhi, N.; Armstrong, L. J.; Petkar, O.; Tripathy, A. K.** (2016): Rice crop yield prediction in india using support vector machines. *IEEE the 13th International Joint Conference on Computer Science and Software Engineering*.

**Garg, A; Garg, B.** (2017): A robust and novel regression based fuzzy time series algorithm for prediction of rice yield. *International Conference on Intelligent Communication and Computational Techniques (ICCT) Manipal University Jaipur*.

**Geeta, M. C. S.** (2015): A survey on data mining techniques in agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no. 2, pp. 887-892.

**Holzman, M. E.; Rivas, R. E.** (2016): Early maize yield forecasting from remotely sensed temperature/vegetation index measurements. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 507-519.

**Iordache, M. D.; Tits, L.; Bioucas-Dias, J. M.; Iaza, A.; Somers, B.** (2014): A dynamic unmixing framework for plant production system monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2016-2034.

**Irwin, S. H.; Sanders, D. R.; Good, D. L.** (2014): Evaluation of selected USDA WAOB and NASS forecasts and estimates in corn and soybeans.

<https://ageconsearch.umn.edu/record/183477>.

**Kaur, M.; Gulati, H.; Kundra, H.** (2014): Data mining in agriculture on crop price prediction: techniques and applications. *International Journal of Computer Applications*, vol. 99, no. 12, pp. 1- 3.

**Medar, R.; Rajpurohit, V.** (2014): A survey on data mining techniques for crop yield prediction. *International Journal of Advance Research in Computer Science and Management Studies*, vol. 2, no. 9, pp. 59-64.

**Meroni, M.; Fasbender, D.; Balaghi, R.; Dali, M.; Leo, O.** (2016): Evaluating NDVI data continuity between SPOT-VEGETATION and PROBA-V missions for operational yield forecasting in north african countries. *IEEE Transactions on Geo-science And Remote Sensing*, vol. 54, no. 2, pp. 795.

**Ramesh, D.; VardhanM, B.** (2015): Analysis of crop yield prediction using data mining techniques. *International Journal of Research in Engineering and Technology*, vol. 4, no.1, pp. 47-473.

**Raorane, A. A.; Kulkarni, R. V.** (2013): Review-role of data mining in agriculture. *International Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 270-272.

**Sharifi, M. A.** (2010): Crop inventory and production forecasting using remote sensing and agrometeorological models: the case of major agricultural commodities in hamadan province, Iran. *International Archives of Photogrammetry and Remote Sensing*, vol. 33, pp. B7.

**Sung, W. T.; Chung, H. Y.; Chang, K. Y.** (2014): Agricultural monitoring system based on ant colony algorithm with centre data aggregation. *IET Communications*, vol. 8, no 7, pp. 1132-1140.

**Tits, L.; Somers, B.; Coppin, P.** (2012): The potential and limitations of a clustering approach for the improved efficiency of multiple end member spectral mixture analysis in plant production system monitoring. *IEEE Transactions on Geo science and Remote Sensing*, vol. 50, no. 6, pp. 2273-2286.

**Wang, Z.; Liang, L.; Slobodan, V.** (2012): Mixture model for multiple instance regression and applications in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp, 2226-2237.