



## Analysis and Prediction of Crop Production in Andhra Region Using Deep Convolutional Regression Network

**Vamsidhar Talasila<sup>1\*</sup>    Chitturi Prasad<sup>1</sup>    Guttikonda Trinesh Sagar Reddy<sup>2</sup>    Allada Aparna<sup>2</sup>**

<sup>1</sup>*Koneru Lakshmaiah Education Foundation, Department of Computer Science and Engineering, India*

<sup>2</sup>*Gudlavalleru Engineering College, Department of Computer Science and Engineering, India*

\* Corresponding author's Email: talasila.vamsi@kluniversity.in

---

**Abstract:** Agriculture planning plays a significant role in economic growth and the food security of agro-based country. Crop yield prediction and selection of crops are the most challenging tasks in agricultural domain and it depends on different parameters such as production rate, market price and government policies. Among the two primary tasks, the crop yield prediction is one of the most demanding tasks for every nation. Due to uncertain climatic changes, farmers are struggling to attain a satisfactory amount of yield from the crops. Many researchers have studied on the prediction of weather, prediction of yield rate of crop, crop classification and soil classification for agriculture planning using statistical methods or machine learning techniques. This study focuses on the prediction of major crops in Andhra Pradesh region and presents an enhanced algorithm known as Deep Convolutional Regression Network (DCRN), which is trained and tested on agricultural data collected from farmers. The experimental results showed that the DCRN method achieved nearly 97% prediction accuracy when compared with existing methods like Decision Tree (DT), Self-Organizing Map (SOM).

**Keywords:** Agricultural, Climatic changes, Crop selection, Crop yield prediction, Machine learning techniques, Soil classification.

---

### 1. Introduction

Agriculture is the backbone of the Indian economy as roughly 70-75% of the population depends directly or indirectly on agriculture. The India's economic growth is directly proportional to agriculture industry growth. However, Indian farmers are facing different challenges due to several constraints like changing climatic conditions, unhealthy soil conditions, soil erosion, land drifting, floods and storms [1]. If agriculture fails, it impacts heavily on the productivity of other industries which in turn impacts on the gross domestic product ratio of the country [2]. The agricultural growth has been stagnant in India and needs a technological revolution which can help in to meet the growing population requirements of the country. In order to feed the increasing population of India, there is a need to incorporate the latest technological methodologies in the agricultural sector. In addition

to this, farmers require timely advice to predict the crop productivity so that they can make proper strategies to increase the yield of their crops [3]. The researchers used precision agriculture techniques to make the soil and crops for optimum productivity and health. In precision agriculture, the real-time farm and weather data are collected using the sensors and the collected data are used to help farmers in taking correct farm-related decision [4, 5]. Sensors are deployed in the farmer fields, where it collects and sends the agricultural data to the relevant data storage node. As, the collected data are huge in volume, hence it can be processed using big data analytics. Big data provide facilities like data storage, data processing and data analysis with accuracy, hence farmers gained benefits in the agricultural fields, which leads to increase in the nation's economic growth [6, 7]. With the help of big data analytics and related Machine Learning

(ML) algorithms, crop productivity can be increased by many folds [8].

Agricultural researchers are exploring a healthy yield prediction based on the group of agricultural data and developing crop yield forecasting methods to increase agricultural and rural statistics [9]. The benchmark dataset is not available for agricultural research and it differs based on different locations, type of crop, climatic condition and irrigation methods [10, 11]. Researchers are using data driven models to get an accurate prediction by utilizing the available data. With data driven model, the ML algorithms play an important role in attaining better accuracy [12, 13]. Though, ML has gained much attention in many areas, the ML techniques still have some restrictions when it is used in a purely data-driven manner. The correctness of the predictions and their reservations developed by the ML algorithms depend on the data quality, model representativeness and the reliance between the input and target variables in the collected datasets [14, 15]. Data with high level of noise, erroneous data, presence of outliers, biases and incomplete datasets may significantly reduce the predictive efficiency of the models [16, 17]. To overcome the issues, this research paper designed a DCRN model to predict the crop yield by using rainfall parameter. The objective of this research is to establish a relationship between summer monsoon variability and shifts of crop production over India for long-range prediction. The major contribution of the research work is to predict the rainfall by using DCRN with minimal set of meteorological parameters of collected data from Andhra Pradesh, India. The proposed method tried to predict higher crop yields, which is used to increase crop intensity, employment, stability and also reduced emigration. The comparative evaluation of DCRN with existing different parametric and non-parametric models is done to validate the adequacy of DCRN in rainfall prediction.

This research paper is organized as follows, Section 2 describes the review of existing techniques, which are used to improve crop productivity. The proposed DCRN methodology is used to predict the crop productivity are represented in Section 3. The experimental results and the validation of the DCRN method is depicted in Section 4. Finally, the conclusion of the paper with future direction is represented in Section 5.

## 2. Literature Review

In this section, the discussion of recent techniques, used to predict crop productivity in

various regions are presented. The advantages and limitations of existing techniques are also depicted in this section.

H. Aghighi, M. Azadbakht, D. Ashourloo, H. S. Shahrabi, and S. Radiom, [18] designed several machine learning techniques, namely Boosted Regression Tree (BRT), Random Forest Regression (RFR), Support Vector Regression (SVR), and Gaussian Process Regression (GPR) approaches to investigate the silage maize yield prediction. The experiments were carried out on time series of Normalized Difference Vegetation Index (NDVI) dataset, which was derived from Landsat 8 OLI images. The results stated that the BRT, GPR and RFR achieved higher performance compared to conventional regression methods because those techniques deal with high-dimensional data of complex distributions as well as the inconsistency of NDVI time series. However, this approach didn't consider the effect of other environmental parameters: soil moisture, texture of soil, climatic parameters and so on.

L. Zhang, J. Jia, G. Gui, X. Hao, W. Gao, and M. Wang, [19] implemented the improved deep learning based classification method to identify the tomato maturity levels according to storage time and appearance. To validate the model on various augmented datasets, a classification system based on Convolutional Neural Network (CNN) was designed for choosing an optimal dataset. The experiments were carried out on different groups of datasets, where best results were predicted by training those datasets. The over-fitting problem was avoided and the CNN method achieved 91.9% accuracy compared to other existing techniques. This proposed method obtained a better result by adding the three major noises, such as Salt, pepper and Gaussian to the dataset. If these noises were not introduced in the dataset, the deep learning techniques provide poor performance.

E. Khosla, R. Dharavath, and R. Priya, [20] designed a Modular Artificial Neural Network (MANN) for predicting the rainfall amount to improve the yield of major kharif crops in Visakhapatnam. The important features were selected by using SVR to identify the various types of kharif crops, bajra, rice, ragi and maize. The experiments were conducted on collected data of recent year and compared it with crop data of the year 1997. The MANN-SVR method used only on rainfall and area attribute in predicting the yield of crops, but the yield of the certain crop depends on many other factors like fertilizers used, irrigation and many more.

B. R. Parida, and A. K. Ranjan, [21] used the remote-sensing (RS) based NDVI profile to identify the wheat crop by deriving the rules for DT classifier. The key findings indicated that the acreage of wheat was estimated and also the long-term wheat statistics data were used to derive a yield model. The DT approach used to produce the acreage map based on the usages of the multi-temporal data (Landsat-8) which was strictly nonparametric and based on the expert knowledge. The experimental study stated that the DT was reliable, cost-effective, and time-saving technique, which was used to make a decision by giving an early estimation of yield. The method didn't focus on other crops and parameters, which affects the yield of wheat crop.

P. M. Gopal, and R. Bhargavi, [22] proposed the hybrid Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) to predict the accurate crop yield. MLR intercept and coefficients were applied to initialize the ANN's input layer bias and weights. The results showed that the prediction accuracy was increased and obtained less number of optimal errors. The prediction accuracy was calculated by using standard performance metrics. The results were compared with traditional techniques including conventional statistical model MLR, conventional ANN, SVR, KNN and random forest. The results showed that the proposed hybrid MLRANN model gave a better prediction accuracy than other models of the same agricultural dataset. The production value decrease, when the maximum temperature increase beyond the threshold value.

J. R. Lamichhane, J. Constantin, C. Schoving, P. Maury, P. Debaeke, J. N. Aubertot, and C. Durr, [23] conducted field experiments in order to improve the seed germination and seedling growth in relation to water content, soil structure, temperatures. Then, this method used the available data to parameterize the SIMPLE crop emergence model and to evaluate its prediction quality. The experimental results showed that soybean germination was very fast, taking only 17 days to reach 50% germination at optimal conditions. The SIMPLE model finely predicted germination and emergence courses and their final rates, compared with the observed field data. The major sources of the non-emergence were: seedling mortality due to soil surface crust, non-germination, and drought. However, the results indicated that when sown in mid-April, seedling mortality due to drought was greater compared with earlier sowing dates.

P. Mohan, and K. K. Patil, [27] implemented an efficient dimensionality technique SOM with Latent

Dirichlet Allocation (LDA). The climate was accurately predicted by using the dimensionality reduced features, which was given as an input to Deep Neural Network (DNN). The experiments were carried out on collected data from Indian meteorological data and Raithamitra-Karnataka State Department of Agriculture (KSDA) by means of precision, recall, accuracy, specificity and sensitivity. However, the method provided poor performance on large-scale collected data in a day wise prediction. Q. Ouyang, and W. Lu, [28] predicted the monthly rainfall by using two novel methods such as Multi-Gene Genetic Programming (MGGP) and Echo State Networks (ESN). The raw rainfall data were decomposed by pre-processing techniques, namely singular spectrum analysis, wavelet transform and ensemble empirical mode decomposition. The experiments were conducted on three rainfall stations in western Jilin Province, China in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Compared to the existing SVR technique, those techniques achieved better performance for month-wise rainfall prediction. However, the risk of climate change was not considered in this study, which lead to poor performance.

### 3. Proposed methodology

In agricultural planning process, agroeconomists need simple and precise estimation techniques to predict yields. Crop yield prediction depends on input features such as area, irrigation methods, temperature and crop yield prediction algorithms. The accuracy of crop yield can be achieved by adopting proper inputs and models without disturbing the agricultural production nature and systems. But, there are some limitations presents in the agriculture planning process. The education and the technical awareness of the farmers in developing countries like India are insufficient to handle modern developments happening around the western countries [24]. Nowadays, the farmers are unaware of the innovative cultivation schemes, tools and technologies. Therefore, they are unaware of genetically engineered seeds that drastically improve yield after harvest.

There is a serious need to promote the technical ability of the farmers to make them competent with the ongoing technological developments happening around the world. The success of agricultural production mainly depends on selecting the appropriate crop suitable for the agricultural land [25].

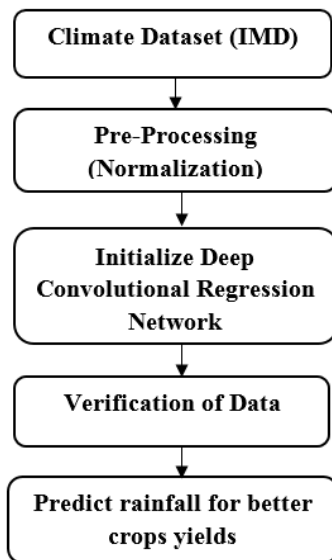


Figure. 1 System model for crop production

Decision on crop selection cannot be done by considering one factor, but many factors need to be considered before taking a decision. Multi-criteria Decision Making methods (MCDM) are used in taking decisions where multiple criteria are considered.

In order to take the proper decisions for crop selection, the agriculture decision making methods considered different criteria, because the various criteria contain different degrees of importance. In this research work, a Deep Convolutional Regression Network (DCRN) is used to predict the major parameters for improving crop productivity. Fig. 1 shows the system model for prediction.

### 3.1 Crop classification

An agro-based country depends on agriculture for their economic growth. When the population of the country increases the dependency on agriculture also increases and subsequent economic growth of the country is affected. In this situation, the crop yield rate plays a significant role in the economic growth of the country. Therefore, there is a need to increase crop yield rate. Crop production rate mainly depends on the geography of a region (e.g. hill area, river ground, depth region), weather condition (e.g. temperature, cloud, rainfall, humidity), soil type (e.g. sandy, silty, clay, peaty, saline soil), soil composition (e.g. PH value, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron) and harvesting methods. Some biological approaches (e.g. seed quality of crop, crop hybridization) and some chemical approaches (e.g. use of fertilizer, urea, potash) are carried out to solve this issue. In addition

to these approaches, a crop sequencing technique is required to improve the net yield rate of crop over the season. This paper proposed a method to achieve the net yield rate of crops over a season. The crop can be classified as:

- a) Seasonal crops that can be planted during a season. e.g. wheat, cotton.
- b) Whole year crops: crops can be planted during the entire year. e.g. vegetable, paddy, Toor.
- c) Short time plantation crops: that take a short time for growing. e.g. potato, vegetables, ratois.
- d) Long-time plantation crops: These crops take a long time for growing. e.g. sugarcane, canda.

A combination of these crops can be selected in a sequence based on yield rate per day.

### 3.2 Dataset description and pre-processing

Initially, input data are taken for experimental analysis from the Indian Meteorological Department (IMD) database, where various datasets are collected and created from a different source. The major regions of Andhra Pradesh are used as an input data to predict the rainfall and improve the crop productivity. The input data is pre-processed to remove the noises and the missing data. Several data pre-processing methods are utilized such as Data Cleaning, Normalization, Discretization, Transformation, and Integration of data. Here, normalization method is used for pre-processing the data to reduce the dimensionality reduction. The input data are taken from the database and data consists of different kinds of units as the temperature is Celsius, wind speed is miles per hour, etc. However, using the normalization method makes the multiple data as single data.

Once the data has been extracted and divided into training, test, and validation, then the method normalized variables to a [0; 1] interval in order to avoid the effects of scale in deep learning architecture. The normalization method employed for the dataset can be observed in the Eq. (1).

$$V_i = \frac{a_i - \min_{a_i}}{\max_{a_i} - \min_{a_i}} \quad (1)$$

Where,  $a_i$  represents a value to normalize for the  $i^{th}$  variable,  $\min_{a_i}$  is the minimum valor registered for this variable in the training set and  $\max_{a_i}$  is the maximum valor registered for this variable in the training set.

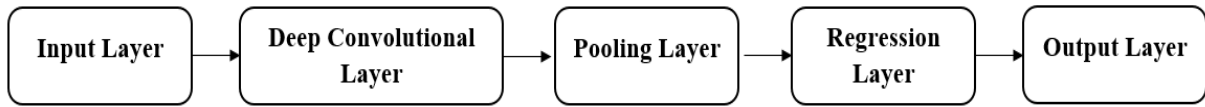


Figure. 2 General architecture of DCRN classifier

The rainfall prediction is considered as one of the important parameters to predict the crop production. Therefore, in this proposed study, DCRN is used to predict the rainfall for improving the crop productivity.

### 3.3 Deep convolutional regression network

The convolutional layer is a primary layer in CNN classifier, which extracts the local information of the data. Moreover, convolutional operation improves the input features and reduce noise interference. Fig. 2 represents the general architecture of DCRN classifier.

The mapping operation in the convolution process is mathematically expressed in Eq. (2).

$$x_j^l = f_c \left( \sum_{i \in M_j} x_i^{l-1} \times k_{i,j}^l + \theta_j^l \right) \quad (2)$$

Where,  $x_j^l$  is specified as the  $j^{th}$  mapping set of convolutional layer  $l$ ,  $x_i^{l-1}$  is denoted as the  $i^{th}$  feature set indicating in the  $(l - 1)$  convolutional layer, and  $k_{i,j}^l$  is indicated as the convolutional kernel between the  $i^{th}$  feature set and  $j^{th}$  mapping set in the convolutional layer  $l$ . The variable  $\theta_j^l$  is represented as bias and  $f_c$  is denoted as activation function. The next step is the pooling process, which reduce the possibility of over-fitting during the training process. The pooling process is mathematically denoted in Eq. (3).

$$x_j^l = f_p(\beta_j^l \text{down}(x_i^{l-1}) + \theta_j^l) \quad (3)$$

Where,  $\text{down}(\cdot)$  is represented as the downsampling approach from layer  $(l - 1)$  to layer  $l^{th}$ ,  $\theta_j^l$  and  $\beta_j^l$  are indicated as the additive bias and multiplicative bias, and  $f_p(\cdot)$  is represented as the activation function. Generally, the pooling process is sub-divided into two types such as, average and maximum pooling. The final pooling layer (matrix features) are arranged to form a rasterization layer, which is further connected with the fully connected layer. The output of node  $j$  is mathematically stated in the Eq. (4).

$$h_j = f_h \left( \sum_{i=0}^{n-1} w_{i,j} x_i - \theta_j \right) \quad (4)$$

Where,  $w_{i,j}$  is denoted as the connection weight of input vector  $x_i$ ,  $\theta_j$  is stated as the node threshold, and  $f_h(\cdot)$  is represented as the activation function.

The next layer is the regression layer that helps to capture the sequential data by considering the prior data. This layer considers the output vectors from the pooling layer as inputs. The regression layer has a number of cells or units and the input of every cell is the output from the pooling layer.

The general architecture of regression method is described in this paper, which employs a deep learning architecture to predict the accumulated rainfall for the next day. The architecture is composed of two networks: An Auto-Encoder network (AEN) and a Multi-Layer Perceptron (MLP) network. The AEN is a deep learning technique which is responsible for feature selection and treatment in time series. An MLP network is responsible for classification and prediction of task. Fig. 3 represents the architecture of the regression method [26].

The first element in DRN architecture is the AEN, an unsupervised network that aims to extract non-linear features for a data input. Being more specific, an AEN is composed of three layers: the input layer, a hidden layer using the sigmoid activation function, and the output layer. Compared to the other classic neural networks, AEN are differently trained so that the output layer attempts to be as similar as possible to the input layer. This way, the number of mixed layer nodes should be lesser than input nodes to achieve dimensionality reduction and uniform mixture.

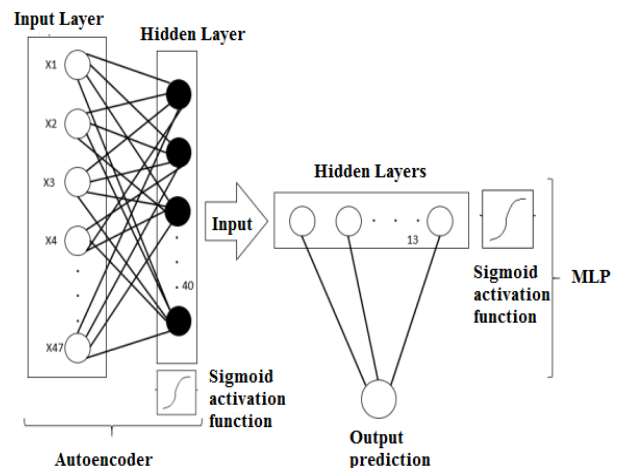


Figure. 3 Architecture of regression

In this network, hyperbolic tangent transfer (tansig) activation function is used in the first deep hidden layer. The second hidden layer log-sigmoid (logsig) activation function is used to increase the correlation within the target data from the Non-linear embedding node. The following activation functions are used in the DRN layers and are given below in Eqs. (5) and (6).

$$\text{Logsig}(n) = \frac{1}{1+e^{-n}} \quad (5)$$

$$\text{Tansig}(n) = \frac{2}{1+e^{-2n}} - 1 \quad (6)$$

The training of the network is done by Levenberg-Marquardt (LM) back propagation technique. The LM algorithm uses second order training speed. For the sum of squares performance function the Hessian matrix can be approximated as  $H = J^T J$  and the gradient can be computed as  $g = J^T e$ , where  $J$  is the Jacobean matrix that contains first derivatives of the network errors with respect to the weights and biases, and  $e$  is a vector of network errors. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update in Eq. (7).

$$X_{k+1} = X_k [J^T J + \mu I]^{-1} J^T e \quad (7)$$

$\mu$  decrease after each step and increased whenever there is an increase in the performance function.

Rainfall data are extracted for the data storage by means of a query, and then the information is normalized by using AEN. The auto-encoder builds the compact non-linear representation of the rainfall

data and then it is fed to the MLP, which is responsible for making the predictions.

## 4. Results and discussion

This section detailed about the experimental result and discussion of the proposed system and also detailed the performance metric, experimental setup, quantitative analysis and comparative analysis. The proposed system was implemented using Python with 4 GB RAM, 1 TB hard disk, 3.0 GHz Intel i5 processor. The performance of the proposed system was compared with other classification methods in order to assess the effectiveness of the proposed system. The performance of a proposed system was evaluated in terms of precision, recall, classification accuracy and f-measure.

### 4.1 Performance measure

The performance measure is defined as the regular measurement of outcomes and results that develops reliable information about the effectiveness of the proposed system. Also, the performance measure is the process of reporting, collecting and analysing information about the performance of a group or individual. The mathematical equation of accuracy, f-measure, precision, and recall are denoted in the Eqs. (8), (9), (10), and (11).

$$\text{Accuracy} = \frac{TN+TP}{TP+TN+FN+FP} \times 100 \quad (8)$$

$$F - \text{measure} = \frac{2TP}{(2TP+FP+FN)} \times 100 \quad (9)$$

$$\text{Precision} = \frac{TP}{(FP+TP)} \times 100 \quad (10)$$

Table 1. Sample data for Andhra region crops

Data Samples	Cloud Cover	MinTemp	AvgTemp	MaxTemp	Precipitation	Vapor Pressure	Rainfall	y_test	predictions	Crop_expert	Crop_Predicted
0	10.063	16.626	21.756	26.898	0	18.923	0	0	0	Rice	Rice
1	14.538	19.608	24.733	29.874	0	21.37	0	0	0	Rice	Rice
2	16.185	17.775	22.939	28.104	0	19.629	0.1	0	0	Rice	Sugarcane
3	16.632	17.444	22.773	28.164	0	19.583	0.2	0	0	Rice	Rice
4	18.407	17.461	22.59	27.732	0	18.534	0	0	0	Rice	Rice
5	11.62	17.208	22.283	27.385	0.847	19.127	1.5	0	0	Rice	Rice
6	23.886	19.117	24.102	29.134	7.065	20.185	9.1	0	0	Onion	Rice
7	24.425	19.289	24.419	29.56	0.15	20.185	0.6	0	0	Rice	Rice
8	16.652	19.89	25.028	30.17	2.466	21.19	0.5	0	0	Rice	Rice
9	19.057	18.712	23.502	28.309	23.541	19.952	16	0	1	Rice	Sugarcane
10	18.166	19.034	24.135	29.26	0.707	20.628	1.6	0	0	Rice	Rice
11	17.932	18.951	24.073	29.214	0.034	20.185	0.1	0	0	Rice	Onion
12	17.932	19.644	24.781	29.921	8.014	21.312	5.7	0	0	Rice	Rice

Table 2. Performance of proposed DCRN method

Crops	Accuracy (%)	F-Measure (%)	Precision (%)	Recall (%)
Rice	97.89	96.35	97.65	95.25
Sugarcane	96.45	95.35	96.12	96.64
Onion	94.32	94.36	94.27	94.57

$$Recall = \frac{TP}{(FN+TP)} \times 100 \tag{11}$$

Where, *TP* is signified as true positive, *TN* is indicated as true negative, *FP* is specified as false positive, and *FN* is indicated as false negative. In Andhra region, the crop rice, sugarcane and onion are majorly sown at low, medium and high rainfall respectively. The input data is collected from the IMD for Andhra regions, where Table 1 shows the sample data for monthly average rainfall of one year.

The data collected for nearly 20 years are taken from the Andhra region, the rainfall values can be calculated by using predicted values from the sample table. The values 0 in a predicted column indicate the low rainfall, whereas 1 indicates medium rainfall and finally 2 represents the high rainfall values.

#### 4.2 Performance analysis of DCRN method

In this section, the proposed DCRN method is validated by using the parameters such as accuracy precision, recall and f-measure. The proposed method is considered the most cultivated crops, namely rice, sugarcane and onion in Andhra Region.

Table 2 shows the validated results of DCRN method and analysis of proposed DCRN by means of various parameters are given in Fig. 4.

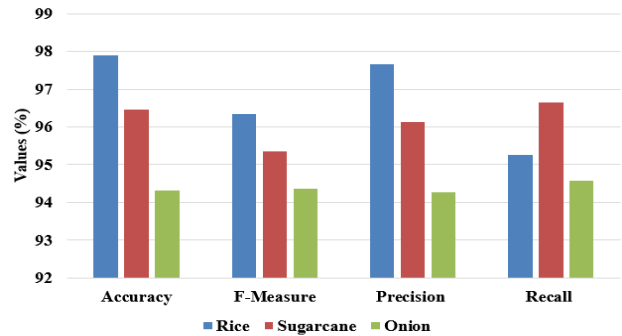


Figure.4 Analysis of Proposed DCRN by means of various parameters

The DCRN method achieved an accuracy of 97.89% for rice, 96.45% accuracy for sugarcane and 94.32% accuracy for onion crop. In addition, the DCRN method achieved nearly 95% of precision, recall and f-measure for onion crops, where 96% of recall, precision and f-measure for sugarcane crop.

From the above results it can be proved that proposed DCRN method achieved better results in various performance measures such as accuracy, recall, precision and F-Measure for major crops of Andhra Region.

#### 4.3 Comparative analysis

The forecasted values of rainfall obtained in this study can be used in agricultural and water resource planning, hydrological model study and climate change study. Table 3 shows the performance criteria of different forecasting methods for rainfall forecasting. The proposed DRN is compared with the existing techniques such as hybrid techniques namely BRT, RFR, SVR and GPR [18], DT [21], Deep Learning based Weighted SOM [27] and ESN, SVR, and MGGP [28] and evaluated in the combinations of testing and training percentage like 80% training and 20% testing of collected data.

Table 3. Comparison of proposed method with existing methods

Authors	Methodology Used	RMSE	MAE	Accuracy
H. Aghighi, M. Azadbakht, D. Ashourloo, H. S. Shahrabi, and S. Radiom [18]	Hybrid Techniques	6.195	4.64	Not Available
B. R. Parida, and A. K. Ranjan [21]	DT	Not Available	Not Available	80%
P. Mohan, and K. K. Patil, [27]	Weighted SOM	Not Available	Not Available	78.98%
Q. Ouyang and W. Lu, [28]	SVR	5.917	4.593	Not Available
	MGGP	7.024	5.480	Not Available
	ESN	3.341	1.890	Not Available
Proposed Method	DCRN	2.148	1.597	97.89%

The existing methods mainly focused on predicting rainfall directly without extracting the useful information from unstructured data. Hence, the performance of these methods provides poor performance in terms of less accuracy. But, the proposed DCRN method focused to extract the important features for predicting the rainfall amount by using effective convolutional neural networks. Features such as cloud cover, average temperature, crop selection, etc. are extracted from the unstructured data to predict the major crop productivity in the Andhra Region. From the above results, the proposed DCRN method achieved nearly 97% of accuracy compared to the other techniques.

## 5. Conclusion

At the present time, one of the most important sources of survival as well as the most crucial factor in the growth of Indian economy is agriculture. More than 70% of the Indian population are involved in agricultural activities. Crop growth and yield models are based on a combination of soil, crop, and climatic variables. The evaluation of the yield in rain-fed agriculture is frequently affected by the water unavailability. The significant factors in crop yield models are climatic information and multiple soil properties are related to crop rooting depth and water availability. This paper evaluated a statistical investigation of rainfall forecasting that was conducted in the Andhra region. The scope of this experiment was to enhance the major crop productivity in high rainfall regions by examining accurate rainfall and crop prediction. In this scenario, a DCRN method was implemented in order to predict the suitable major crop for the season. The evaluation result of the proposed method showed better performance than the existing methods. On average, the proposed method achieved 97.89% accuracy by utilizing a DCRN algorithm. Related to the other approaches for crop prediction, the advanced scheme delivered an effective performance by means of accuracy, precision, error rate and F-Measures than the previous methods. In the future, for enhancing the agricultural productivity, the crop prediction can be further improved by employing a new approach along with a number of other major factors such as soil, risk and so on.

## References

[1] C. Zhang, L. Di, L. Lin, and L. Guo, "Machine-learned prediction of annual crop planting in the US Corn Belt based on historical crop planting

- maps", *Computers and Electronics in Agriculture*, Vol.166, pp.104989, 2019.
- [2] S. Shidnal, M.V. Latte, and A. Kapoor, "Crop yield prediction: two-tiered machine learning model approach", *International Journal of Information Technology*, pp.1-9, 2019.
- [3] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks", *Computers and Electronics in Agriculture*, Vol.163, pp.104859, 2019.
- [4] W.D.S.G. Júnnyor, E. Diserens, I.C. De Maria, C.F. Araujo-Junior, C.V.V. Farhate, and Z.M. de Souza, "Prediction of soil stresses and compaction due to agricultural machines in sugarcane cultivation systems with and without crop rotation", *Science of The Total Environment*, Vol.681, pp.424-434, 2019.
- [5] A. Haghverdi, R.A. Washington-Allen, and B. G. Leib, "Prediction of cotton lint yield from phenology of crop indices using artificial neural networks", *Computers and Electronics in Agriculture*, Vol.152, pp.186-197, 2018.
- [6] A. Kamilaris, A. Kartakoullis, and F.X. Prenafeta-Boldú. "A review on the practice of big data analysis in agriculture", *Computers and Electronics in Agriculture*, Vol.143, pp.23-37, 2017.
- [7] L. He, C.A. Coburn, Z.J. Wang, W. Feng, and T.C. Guo, "Reduced Prediction Saturation and View Effects for Estimating the Leaf Area Index of Winter Wheat", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.57, No.3, 1637-1652, 2018.
- [8] A. Anitha, and D.P. Acharjya, "Crop suitability prediction in Vellore District using rough set on fuzzy approximation space and neural network", *Neural Computing and Applications*, Vol.30, No.12, pp.3633-3650, 2018.
- [9] B. Basso, D. Cammarano, and E. Carfagna, "Review of crop yield forecasting methods and early warning systems", In: *Proc. of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics*, FAO Headquarters, Rome, Italy, Vol.41, 2013.
- [10] W. Lu, David E. Atkinson, and N.K. Newlands. "ENSO climate risk: predicting crop yield variability and coherence using cluster-based PCA", *Modeling Earth Systems and Environment* Vol.3, No.4, pp.1343-1359, 2017.
- [11] D. Helman, I.M. Lensky, and D.J. Bonfil, "Early prediction of wheat grain yield production from root-zone soil water content at heading using Crop RS-Met", *Field Crops Research*, Vol.232, pp.11-23, 2019.



- [12] L. Xing, L. Li, J. Gong, C. Ren, J. Liu, and H. Chen, "Daily soil temperatures predictions for various climates in United States using data-driven model", *Energy*, Vol.160, pp.430-440, 2018.
- [13] L. Hund, B. Schroeder, K. Rumsey, and G. Huerta, "Distinguishing between model-and data-driven inferences for high reliability statistical predictions", *Reliability Engineering & System Safety*, Vol.180, pp.201-210, 2018.
- [14] S. Liu, X. Wang, M. Liu, and J. Zhu, "Towards better analysis of machine learning models: A visual analytics perspective", *Visual Informatics*, Vol.1, No.1, pp.48-56, 2017.
- [15] Y.H. Kuo, Z. Li, and D. Kifer, "Detecting Outliers in Data with Correlated Measures", In: *Proc. of the 27<sup>th</sup> ACM International Conference on Information and Knowledge Management*. ACM, pp.287-296, 2018.
- [16] K. Gibert, M. Sánchez-Marrè, and J. Izquierdo. "A survey on pre-processing techniques: Relevant issues in the context of environmental data mining", *AI Communications*, Vol.29, No.6 pp.627-663, 2016.
- [17] S. K. Kwak, and J. H. Kim, "Statistical data preparation: management of missing values and outliers", *Korean Journal of Anesthesiology*, Vol.70, No.4, pp.407, 2017.
- [18] H. Aghighi, M. Azadbakht, D. Ashourloo, H. S. Shahrabi, and S. Radiom, "Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol.11, No.12, pp.4563-4577, 2018.
- [19] L. Zhang, J. Jia, G. Gui, X. Hao, W. Gao, and M. Wang, "Deep learning based improved classification system for designing tomato harvesting robot", *IEEE Access*, Vol.6, pp.67940-67950, 2018.
- [20] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression", *Environment, Development and Sustainability*, pp.1-22, 2019.
- [21] B. R. Parida, and A. K. Ranjan, "Wheat Acreage Mapping and Yield Prediction Using Landsat-8 OLI Satellite Data: a Case Study in Sahibganj Province, Jharkhand (India)", *Remote Sensing in Earth Systems Sciences*, Vol.2, No.2-3, pp.96-107, 2019.
- [22] P.M. Gopal, and R. Bhargavi, "A novel approach for efficient crop yield prediction", *Computers and Electronics in Agriculture*, Vol.165, pp. 104968, 2019.
- [23] J.R. Lamichhane, J. Constantin, C. Schoving, P. Maury, P. Debaeke, J. N. Aubertot, and C. Durr, "Analysis of soybean germination, emergence, and prediction of a possible northward establishment of the crop under climate change", *European Journal of Agronomy*, Vol.113, pp. 125972, 2020.
- [24] P.K. Jha, P. Athanasiadis, S. Gualdi, A. Trabucco, V. Mereu, V. Shelia, and G. Hoogenboom, "Using daily data from seasonal forecasts in dynamic crop models for yield prediction: A case study for rice in Nepal's Terai", *Agricultural and Forest Meteorology*, Vol.265, pp.349-358, 2019.
- [25] A. Khamparia, G. Saini, D. Gupta, A. Khanna, S. Tiwari, and V. H. C. de Albuquerque, "Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network", *Circuits, Systems, and Signal Processing*, pp.1-19, 2019.
- [26] E. Hernandez, V. Sanchez-Anguix, V. Julian, J. Palanca, and N. Duque, "Rainfall prediction: A deep learning approach", In: *Proc. of International Conference on Hybrid Artificial Intelligence Systems*, pp.151-162, 2016.
- [27] P. Mohan and K.K. Patil, "Deep Learning Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application", *International Journal of Intelligent Engineering and Systems*, Vol.11, No.4, pp.167-176, 2018.
- [28] Q. Ouyang and W. Lu, "Monthly Rainfall Forecasting Using Echo State Networks Coupled with Data Preprocessing Methods", *Water Resources Management*, Vol.32, No.2, pp.659-674, 2018.