# Predicting crop yield with NDVI and Backscatter values using Deep Neural Networks

Pragathi Tummala
*Department of CSE*
*V R Siddhartha Engineering College*
Vijayawada, Andhra Pradesh, India
pragathitummala2k1@gmail.com

M Sobhana
*Faculty at Department of CSE*
*V R Siddhartha Engineering College*
Vijayawada, Andhra Pradesh, India
sobhana@vrsiddhartha.ac.in

Sruthi Kakumani
*Department of CSE*
*V R Siddhartha Engineering College*
Vijayawada, Andhra Pradesh, India
sruthikakumani@gmail.com

***Abstract*—Agriculture is India's major cause of income and even plays a critical part in ensuring food security in the Indian economy. India's agricultural sector faces the tough problem of maximizing crop productivity. A common problem for Indian farmers is not selecting the right crops based on their soil needs. This has an impact on productivity. Predicting yield before harvest helps farmers take appropriate trading and repository measures. This work assists agronomists to realize the yield of crops before planting them on their farmland. Consequently, it guides farmers in taking the right decisions. Deep learning (DL) can be an important consideration in obtaining an effective solution for yield forecasting. This study is primarily based on providing accurate yield prediction with a focus on key crops. The deep Neural Network (DNN) model is employed for predicting the crop yield. The proposed deep learning model takes the inputs such as crop, season, Normalized Difference Vegetation Index (NDVI), backscatter values, and area, which are then analyzed in the hidden layers using weights that are adjusted throughout training. Yield prediction of a crop in kilogram (kg) is then spewed out by the model. The findings obtained with this method show higher accuracy.**

***Keywords—Deep Neural Network (DNN), Crop yield, Normalized Difference Vegetation Index (NDVI), Backscatter value, Prediction, Accurate value.***

## I. INTRODUCTION

Agriculture demand rises due to the increase in the global population. Crop phenology's annual cycles and crop production's reliance on climate, soil, and weather conditions make it difficult to govern farming activities. Due to these difficulties, a huge majority of farmers do not make the expected profit. Farmers demand fast advice on crop profitability potential; thus, an analysis must be undertaken to aid farmers in increasing their profits. One of agriculture's most serious concerns is yield estimation. The crop yield is something that every farmer is interested in learning about. It's critical to have adequate knowledge regarding crop production history while making agricultural recommendations. In earlier centuries, crop yield prediction was done by painstakingly assessing the cultivator's prior knowledge of the crop. However, there is a large volume of agricultural data, and manual analysis is laborious. Producers used to predict output by focusing on yields from the previous year. Crop forecast using Machine Learning uses different methodologies or algorithms, and we can anticipate crop yields using such algorithms [1]. Machine learning is now being used in agricultural applications to improve crop yield prediction accuracy. Although machine learning has advanced significantly, its application in data-driven scenarios is still constrained. Its precision is also impacted by the model's representativeness, data quality, and is dependent on the collected dataset's input factors. As a result, techniques of deep learning have been applied for improving the accuracy of agricultural yield prediction. The DNN for agricultural yield prediction represents true data with no handcrafted attributes [2],[3].

Deep Neural Network was created to train nonlinear and intricate relationships between input parameters utilizing past knowledge to get a precise yield forecast from known weather conditions. The DNN is proposed in this research for modelling climatic changes, climate parameters such as crop season, NDVI values, and agricultural production prediction. DNN comprises one input, many hidden, and one output layer. The solution of one layer is passed as an input to another layer and this continues up to the output layer. Each layer consists of several neurons which analyse the given input and produces output. It improves the accuracy of crop yield prediction by taking into account several climate-related data. Based on the existing data about important events and functions related to the outcome, the dynamics that are either lacking or inadequately described in parametric frameworks are captured by augmenting a parametric framework with DNN. The accuracy of DNN is improved by taking into account several climate parameters as well as the effect of weather on agricultural yield.

## II. LITERATURE SURVEY

Saeed Khaki et al., [3] map genotypes, yield performance of approximately 2,300 maize combinations, develop deep neural network (DNN) approaches based on the state-of-the-art model and result approaches, and perform trait selection based on trained DNN models. However, the input space without a succeeded in reducing the significant decrease in prediction accuracy. We have found an improvement in prediction accuracy using the root mean square error.

Vinushi Amaratunga et al., [4] trained a constructed neural network model. Different training approaches like Bayesian regularization, scaled conjugate gradient, and Levenberg Marquardt were studied. The performance of Artificial Neural Network (ANN) models created with MATLAB numerical calculations was assessed. Because the climate parameters are employed as performance measures for the correlation factor (R) and mean square error (MSE) to save computational effort, the Language Model (LM) training approach gives rice yields.

M.K Dharani et al., [5] studied agricultural yield prediction which included a variety of deep learning algorithms, including ANN, Convolutional Neural Network (CNN), DNN, Recurrent Neural Network (RNN), and hybrid networks. Wheat, barley, sugar beet, potato, and sunflower crop pictures were utilized to train ANN. In DNN, the climate is mapped with plant genotype and other plant-growing parameters. The vegetative index and meteorological data were used to monitor the crops in Long Short-Term Memory (LSTM). RNN and hybrid networks outperform the remainder of the network, achieving 90 percent accuracy.

Yuming Guo et al., [6] exercised a diversity of techniques, as well as non-linear Feed-forward backpropagation neural networks (FFBN) tool and linear Partial Least-Squares Regression (PLSR) to estimate rice yields. Variables were discovered among the parameters in the PLSR model, indicating that improvements to climate data-based modelling should be made. A network model with a comparable accuracy rate between both the train and test sets was chosen as the optimal neural network design in FFBN.

Mayank Champaneri et al., [7] built a website and applied machine learning methods to estimate crop productivity before planting on a farm. The random forest approach is used to evaluate various aspects of weather, including temperature, humidity, rainfall, moisture, and so on. It works by training a vast number of decision trees, then generating class output, which is the mean forecast of all the trees or the mode of all the classes.

K Mythili et al., [8] predicted plant cultivation applying the Deep Neural Network of the modified version, then by making use of an optimization technique termed as Particle Swarm Optimization (PSO). Here, L2 regularization is utilized to compute the weight matrix, while PSO is used to change Modified Deep Neural Network (MDNN) hyperparameters and network structure. PSO-MDNN was found to be very accurate at 95.49 percent when compared to Decision Tree, K-Nearest Neighbours (KNN), and Random Forest.

P. Priya et al., [9] forecasted the crop yield using the Random Forest technique. By a 72 percent to 28 percent split ratio, input data is divided into two groups: train and test data. For the needed tuples, the mean and standard deviation were determined. The probability with the highest value is then utilized to construct a prediction based on the data. The accuracy is being evaluated by measuring the generated class label to the validation set. The degree of precision might range from 0% to 100%.

S. Pavani et al., [10] acquired data from several regions of Telangana from the TS Development Society. The K nearest approach, a machine learning technique, was used to forecast crop yield. Various climatic factors, soil necessities were taken into account as closest neighbours for a specific instant. As a result, it was determined that KNN generates a more precise and efficient agricultural yield.

S. Veenadhari et al., [11] anticipated yield using climate information, utilizing a machine learning approach. Crop Advisor, a software program that predicts the impact of climatic conditions on crop yields, was produced as an easy-to-use interface in the present study. C4.5 method creates the most significant weather conditions on-farm yields of preferred crops in regions of Madhya Pradesh. Decision Tree has been used to implement the study.

Subhadra M, Gour et al., [12] used methodologies of Machine Learning in farm management of crops for a variety of applications. An intensified indistinct cluster analysis is offered using GPS- based colour photos for identifying plants, soil, and residue zones of interest. The study comprises several criteria that can aid in improving crop yield and raising the yield ratio during cultivation.

D Ramesh et al., [13] utilized Data Mining methodologies for the analysis of crop yield prediction. The major objective is providing a convenient interface for agriculturalists that provides rice production statistics based on existing data. Various data mining approaches were utilized to estimate crop yield to maximize crop productivity. For ex., K- Means clustering can be applied to predict pollution factors.

Gandhi N et al., [14] combined meteorological and crop yield data from the Kharif season (June-November) during a five-year period (1998-2002) for predicting the yield of rice in some Maharashtra regions. The Bayes classification and Nave Bayes algorithms were utilized. The goal was to study how successful Bayes Net is and compare it to other traditional assessors. It was clear from the outcomes and statistics that Bayes Net outperformed Naive Bayes.

Kaur et al., [15] chose the predictive Apriori algorithm to examine the consequence of everyday heat conditions and daily rain on the yield of paddy in Patiala and Ludhiana areas. The research was carried out using the Waikato Environment for Knowledge Analysis program. Several stages of paddy development were considered in the research.

Khan et al. [16] asserted that numerous association rule mining methods can be used in a variety of agricultural applications. It covers a variety of association algorithms, such as Apriori, Pincer search, and others. These can be used to predict the crop's production.

Bharadi et al., [17] applied data mining approaches including K-means clustering, density-based algorithms, to evaluate agricultural data. The experiment employed a dataset of approximately 4200 instances along with eight features like crop, type, area, state, production, various climatic factors. Their findings revealed rainfall scale from 1306.9mm to 1482.57mm, temperatures range from 24.62°C to 27.183°C, productivity tends to increase.

S. Shanmuganathan et al., [18] did studies to establish a relationship between temperature differences and a grapevine yard in New Zealand. With the help of Neural Networks, the qualities with comparable values were successfully clustered. The relevance of the connected components was then tested using a chi-square test. The findings of the study concluded that temperature variations, in combination with factors such as humidity, wind speed, and precipitation, played a crucial role in determining the yield and quality of the wine.

Md. A. R. Sarker et al., [19] performed regression to determine the relationship between three variables: rice yield, rainfall, and annual temperature. Finally, the amount of annual production is influenced by climate and precipitation, according to the findings.

Narayanan Balkrishnan et al., [20] in their study employed the AdaNaive, AdaSVM approaches as the recommended method for forecasting the production of a crop. Here, AdaNaive and AdaSVM are implemented. Naive Bayes and Support Vector Machine (SVM) algorithms are made more efficient with the AdaBoost approach.

## III. Methodology

The pre-processing of this work uses crop production history and soil needs and is trained by DNN. The DNN architecture consists of input, hidden and output layers; neurons in the hidden layer, an activation function, and a predictive model training approach. DNN is used in classification tasks because of its popularity these days. DNN services are instantly linked to frameworks dynamically designed through network professionals.

### A. *Data Collection*

Deep Learning techniques rely heavily on data. In the Data collection phase, data is collected from multiple sources and a dataset is prepared. Also, the dataset provided will be used by analytics. The data was gathered from different government websites. Various online summaries are available like govt. data and aps.dac.gov.in. Annual extracts from culture are used for at least 10 years. These datasets usually accept chaotic time series behaviour.

### B. *Data Cleaning*

The integrated and developed dataset emerged in various attributes and some of the less important ones were ignored as part of the pre-processing. Therefore, after removing the unnecessary parameters of the predictive effect in this study, we combined various historical records. Parameters crop year, season, crop, area, NDVI, backscatter values were utilized. The data was divided as 70 percent and 30 percent in-place-of training and testing respectively. The classifier was trained and generated to build the model. The model was evaluated for the predictive accuracy of the test data.

### C. *Model creation and Training*

An appropriate activation function is selected for the model. One of the most important components of a neural network is the activation function. Activation introduces a highly desirable non-linearity in the model. The sigmoid activation function has been the preferred choice for many years.

The non-linear sigmoid function is expressed as;

$$f(x) = \frac{1}{(1+e^{-x})} \tag{1}$$

The number of hidden layers for the neural network is chosen for the purpose of training the model. In general, more hidden layers than the average number of layers are good. All regularization methods are at least partially to handle the extra units. On the other hand, it is more likely that the model will not be well-tuned while maintaining the number of hidden units (more than the optimal number). Choosing the optimal number of layers is relatively easy. Keep adding layers until the test error stops improving.

Weights are to be initialized between two neurons in the network. Always initialize weights with a small random number. When the sigmoid function is used as an activation function, if the weight is too high, the sigmoid will reach its tail and the neuron will die. If the weights are very small, the gradient is also small Therefore; it is desirable to choose the weights in the middle range so that they are evenly distributed around the mean.

Learning rate is considered for the process of learning. This is possibly one of the most significant parameters that control the learning process of a neural network. If the training rate is too small, it can take the model to converge forever, and if it is too large, the loss can spike in some early training examples. In general, a learning rate of 0.01 is safe, but it should not be taken as a strict rule. The optimal learning rate is because it needs to correspond to a specific task.

| State_Name | District_Name | Crop_Year | Season | Crop | Yield_kg | Back Scat | NDVI | Area |
|---|---|---|---|---|---|---|---|---|
| Andhra Prad | ANANTAPUR | 2014 | Kharif | Arhar/Tur | 170 | -8.8509 | 0.5594 | 1400 |
| Andhra Prad | ANANTAPUR | 2014 | Kharif | Bajra | 100 | -13.241 | 0.3139 | 1000 |
| Andhra Prad | ANANTAPUR | 2014 | Rabi | Brinjal | 170 | -7.5842 | 0.797 | 7300 |
| Andhra Prad | VISAKHAPATANA | 2014 | Kharif | Cabbage | 40 | -9.1705 | 0.7851 | 3700 |
| Andhra Prad | VISAKHAPATANA | 2014 | Whole Year | Cashewnut | 55 | -9.1304 | 0.7873 | 65080 |
| Andhra Prad | VISAKHAPATANA | 2014 | Whole Year | Castor seed | 50 | -9.2485 | 0.7723 | 3300 |
| Andhra Prad | VISAKHAPATANA | 2014 | Kharif | Coriander | 55 | -11.089 | 0.2357 | 10100 |
| Andhra Prad | WEST GODAVAR | 2014 | Rabi | Cotton(lint) | 95 | -9.5905 | 0.2434 | 2200 |

Fig 1. Snapshot of final dataset used for training the model.

### D. *Predicting Crop Yield Using DNN*

DNN employs a layering system in sequence. The depth of a DNN is determined by the number of layers utilized. In hidden layers of DNN, it uses fuzzy logic. Deep Neural Networks are widely used because they choose the most appropriate characteristics before categorizing them. The proposed DNN [21] classifies the pre-processed data. Decisions about each input node are made by DNN and sent to the next layer for training. Networks use several hidden layers. The weighting and processing of input layer pieces or data points are done by succeeding layers.

During this work we are considering crop season, year, crop, NDVI values, area, backscatter values are taken as input features that are prominent characteristics. These parameters are important for better yield of a crop, even though various aspects are there to consider like temperature, humidity, soil moisture. Many researchers have predicted the crop yield using these factors. We are, for the first time, contemplating these values. NDVI means Normalized Difference Vegetation Index. Values of NDVI scale from -1.0 to +1.0. Barren rocks areas, sandy and snowy areas normally have little values of NDVI, 0.1 and below. Moderate NDVI levels of 0.2-0.5 can be found in scanty vegetation like bushes, meadows, and decrepit plants. Large values of NDVI between 0.6 and 0.9 equate to opaque vegetation as found in mild and equatorial forests and plants at the highest growing stages.

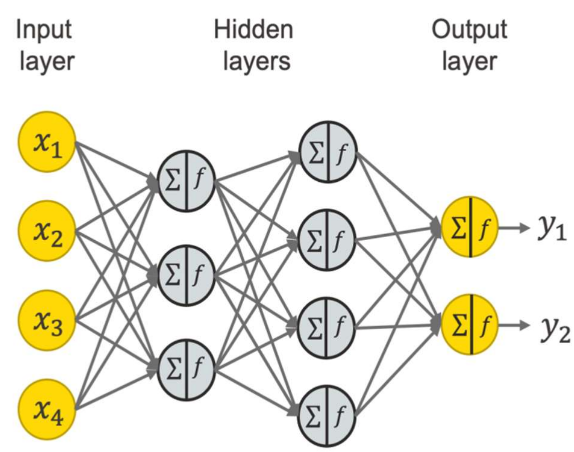


Fig 2. Deep neural network

Backscatter values indicate the roughness levels of the surfaces in the given area due to wind effects and seasonal freezing. Using roughening models, it is feasible to monitor the evaporation rate of the region where the crops are sown quantitatively. The deep neural network is trained with all the above parameters by choosing the appropriate activation function and number of hidden layers, initializing weights, and considering a suitable learning rate for the network. DNN assigns weights to each layer during the process until the final output is obtained. Thus, DNN can predict the crop yield for given parameters. Here we considered a learning rate of 0.01 which is an optimal value that controls the process of learning of a neural network.

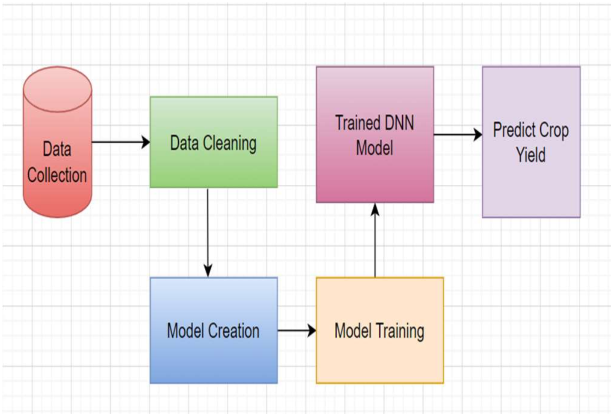


Fig 3. Block diagram of the proposed model

In the case of DNN, the error function E is employed to calculate training accuracy. This procedure of learning is reiterated till DNN evaluates the association among the input, the retrieved data, producing a group of trained categories that form the basis of the prediction.

$$E = \frac{1}{2}(t - y)^2 \tag{2}$$

Where t is the target output and y is the actual output of the neuron.

### *E. Performance Evaluation*

The parameters obtained from the trained model will be examined and accurate prediction values are provided. The model's performance can be evaluated using metrics such as root mean square error, mean absolute percentage error, mean absolute error, mean squared error. Here, a trained model is evaluated with the mean squared error (MSE) metric to assess the loss, and an accuracy of 95.47% is obtained by calculating an average squared difference between the observed and predicted values.

## IV. RESULTS

The model that has been trained can predict crop yield with a low loss factor and high accuracy. The model developed gives the accuracy of 95.47% in the districts of Andhra Pradesh for all the crops during Kharif and Rabi seasons. By using a Deep Neural Network we can predict the crop yield based on several parameters. The model gives the exact matching values of both actual and predicted values. Our model takes the input NDVI, backscatter values, area, season, year, and crop and displays the predicted yield of a crop in Kilogram (kg). The proposed method is helpful for farmers even in their smallest plot of the crop.

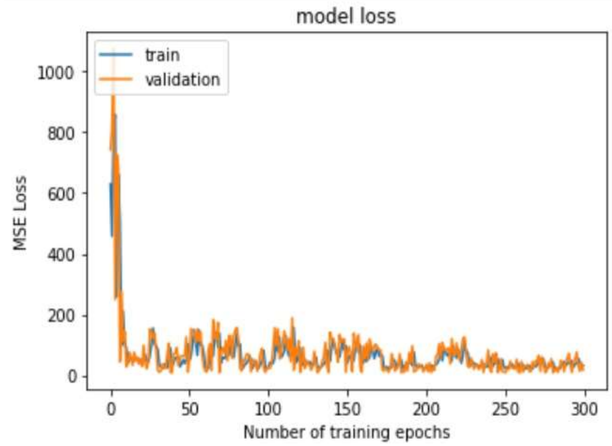


Fig 4. The Plot of Training and validation of loss

## V. CONCLUSION AND FUTURE WORK

Finally, a deep learning technique for estimating agricultural yields was developed, it is effective in predicting crop yield in specific districts of Andhra Pradesh. Most crucially, this research is ahead of the curve since it forecasts agricultural yields. The agriculture sector's throughput is increased with Deep Learning models. Any user can examine which crop type best suits their conditions based on the predicted yield. Using the most accessible technologies, this effort provides direct advising facilities to the farmer in his minimum area. The proposed prediction model can be used by farmers to help them decide what to plant and when to plant it. Future work will focus on developing a system to recommend suitable crops based on various parameters and climatic conditions such as rainfall, temperature, humidity, pH values. The system can be improved to recommend fertilizers to farmers if the nutrient deficiencies in the soil are identified by looking at Nitrogen, Potassium, Phosphorus, and other nutrients of the soil.

## REFERENCES

[1] Rashid, Mamunur, et al. "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction." *IEEE Access* 9 (2021): 63406-63439.

[2] Kalaiarasi, E., and A. Anbarasi. "Crop yield prediction using multi-parametric deep neural networks." *Indian Journal of Science and Technology* 14.2 (2021): 131-140.

[3] Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." *Frontiers in plant science* 10 (2019): 621

[4] Amaratunga, Vinushi, et al. "Artificial neural network to estimate the paddy yield prediction using climatic data." *Mathematical Problems in Engineering* 2020 (2020).

[5] Dharani, M. K., et al. "Review on Crop Prediction Using Deep Learning Techniques." *Journal of Physics: Conference Series*. Vol. 1767. No. 1. IOP Publishing, 2021.

[6] Guo, Yuming, et al. "Prediction of rice yield in East China based on climate and agronomic traits data using artificial neural networks and partial least squares regression." *Agronomy* 11.2 (2021): 282.

[7] Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara, Mansing Rathod,"Cropyield prediction using machine learning" International Journal of Science and Research, April 2020.

[8] Mythili, K., and R. Rangaraj. "Deep Learning with Particle Swarm Based Hyper Parameter Tuning Based Crop Recommendation for Better Crop Yield for Precision Agriculture." *Indian Journal of Science and Technology* 14.17 (2021): 1325-1337.

[9] Priya, P., U. Muthaiah, and M. Balamurugan. "Predicting yield of the crop using machine learning algorithm." *International Journal of Engineering Sciences & Research Technology* 7.1 (2018): 1-7.

[10] Pavani, S., and A. S. Beulet. "Heuristic prediction of crop yield using machine learning technique." *International Journal of Engineering and Advanced Technologies* 9.01 (2019): 135-138.

[11] Veenadhari, S., Bharat Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." *2014 International Conference on Computer Communication and Informatics*. IEEE, 2014.

[12] Mishra, Subhadra, Debahuti Mishra, and Gour Hari Santra. "Applications of machine learning techniques in agricultural crop production: a review paper." *Indian Journal of Science and Technology* 9.38 (2016): 1-14.

[13] Ramesh, D., and B. Vishnu Vardhan. "Analysis of crop yield prediction using data mining techniques." *International Journal of research in engineering and technology* 4.1 (2015): 47-473.

[14] Gandhi, Niketa, Leisa J. Armstrong, and Owaiz Petkar. "PredictingRice crop yield using Bayesian networks." *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2016.

[15] Kaur, Kuljit, and Kanwalpreet Singh Attwal. "Effect of temperature and rainfall on paddy yield using data mining." *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*. IEEE, 2017.

[16] Khan, Farah, and Divakar Singh. "Association rule mining in the field of agriculture: a survey." *International Journal of Scientific and Research Publications* 329 (2014).

[17] Bharadi, Vinayak A., et al. "Analysis and prediction in agricultural data using data mining techniques." *International Journal of Research in Science and Engineering* (2017): 386-393.

[18] Shanmuganathan, Subana, Philip Sallis, and Ajit Narayanan. "Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine, Wine Quality and Perennial Crop Yield." *2010 2nd International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, 2010.

[19] Sarker, Md Abdur Rashid, Khorshed Alam, and Jeff Gow. "Exploring the relationship between climate change and rice yield in Bangladesh: An analysis of time series data." *Agricultural Systems* 112 (2012): 11-16.

[20] Balakrishnan, Narayanan, and Govindarajan Muthukumarasamy. "Crop production-ensemble machine learning model for prediction." *International Journal of Computer Science and Software Engineering* 5.7 (2016): 148.

[21] Deep Neural Network, https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks