International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 8 ǁ 2021 ǁ PP. 52-55

Crop Yield Prediction through different Machine Learning Algorithm

Prabhuprasad Wattamwar [*1], Rutuja Palav[*2], Rushikesh Nilkanthe[*3], Dr Kavita Joshi[*4]

**1 G H Raisoni College of Engineering and Management Pune,*
**2 G H Raisoni College of Engineering and Management Pune,*
**3 G H Raisoni College of Engineering and Management Pune,*
**4 Professor G HRaisoni College of Engineering and Management Pune,*
(Ex- Professor, Department of E&TC Engineering G H Raisoni College of Engineering and Management Pune, maharashtra, India)

Abstract
Precise yield estimation is crucial in agriculture. Remote Sensing (RS) frameworks are extra commonly applied in constructing choice help devices for the modern cultivating frameworks to enhance yield even as diminishing working prices and natural effect. The number of crops inside the area is one of the principal additives for determining crop yield. This counting task is to be done carried out using a human rather than a computer and is hence time-consuming In this paper, we propose a green approach that uses computer vision and precisely count the vegetation in a digital photograph, in addition to we've also devised algorithms which could correctly determine the yield on the idea of given elements along with soil precipitation, Area, humidity Index and greater such elements which performs a crucial position in figuring out the yield. We have used various algorithms such as random forest regression, Support Vector Machine, CNN, and Deep Neural network in this paper and worked on the above problem.

Keywords: *field Prediction, object counting, computer vision, deep learning, convolution neural network, deep neural networks, regression, SVM.*

Date of Submission: 18-07-2021 Date of acceptance: 04-08-2021

I. INTRODUCTION

Farming is a significant source of income for many people in developing countries. In addition, agricultural growth has been more rapid than growth in the non-agricultural sectors in recent years in many countries. The two types of index products are parametric and sample-based. Examples of parametric indices in insurance include weather (with triggers based on variables such as rainfall, temperature, humidity, wind speed, etc.), flooding (water levels and durations triggers), wind speed (velocity and duration triggers), and seismic activity (Richter scale triggers). Test-based records incorporate territory-based yield protection and test-based domesticated animal file protection. Zone yield protection is a choice on the typical yield for creation in a locale/region. On the off chance that the zone is adequately huge, region yield protection is not powerless to moral danger issues since the activities of an individual rancher will have no observable effect on the zone's typical yield. Region yield protection likewise has moderately low exchange costs since there is no compelling reason to set up and check explicit ranch yields for each safeguarded unit, nor is there any need to direct on-ranch misfortune alteration.

II. METHODOLOGY

As this is a regression problem, we have applied three different algorithms, namely Support Vector Regression III-B, Random Forest III-A, Deep Neural Network III-C. In the third dataset, which consists of image data, we have used Convolution Neural Network III-D. Data Preprocessing is done on the CSV data. The Categorical data is encoded through one-hot encoding. Then the data is normalized through the Min-Max scalar, and the missing values in the particular column are replaced with the mean of that column. The python pandas library is used for preprocessing. At last, the data is split into the train(0.7) and test(0.3) before applying the different models. Data preprocessing is done on the image dataset. Since there was no available dataset to download, we had to create our dataset by hand labeling the number of crops per image. We made ten such images for use, which were all different from each other. We had created a CSV file with the image ID and its respective number of crops in the corresponding column. We normalized the varying image sizes into a 400x400

pixel size to feed the images into the model. We split the pictures into eight ideas for training and two doubles for testing. [2]

A. Random Forest Regression

The random forest model is an added substance model that settles on forecasts by joining choices from a succession of base models.

$$g(x) = f0(x)+f1(x)+f2(x) +... (1)$$

where the last model g is the sum of basic base models fi. Here, each base classifier is a decision tree. This expansive procedure of utilizing different models to acquire better prescient execution is called model ensembling. In arbitrary backwoods, all the base models are built autonomously utilizing an alternate sub-test of the information. We have imported Randon Forest Regressor from the sklearn library and computed the given scores for comparison. [3]

B. Support Vector Regression

A support vector machine builds hyperplanes in a high or limitless dimensional space, which can be utilized for arrangement, relapse, or different assignments like exceptions discovery. The Support Vector Regression (SVR) utilizes indistinguishable standards from the SVM for order, with just a couple of minor contrasts. An edge of resistance (epsilon) is set in estimate to the SVM, which would have just mentioned from the issue on account of relapse. However, other than this reality, there is likewise a progressively entangled explanation; the calculation is increasingly confused along these lines to be taken in thought. Notwithstanding, the fundamental idea is consistently the equivalent: to limit mistake, individualizing the hyperplane that augments the edge, remembering that the blunder is endured. We have imported Support Vector Regressor from the sklearn library and computed the given scores for comparison. [3]

C. Deep Neural Network

A deep neural network is a neural network with a certain level of complexity, a neural network with more than two layers. We had imported a sequential neural network from Keras for designing our model. We had used multiple layers of dense, fully connected layers so that Accuracy is high. We have trained the two different neural networks on the two datasets respectively for 50 epochs. The first Neural Network for the first dataset consisted of 1 dense input layer with a relu activation function, three hidden layers with a real activation function, and the output layer was a thick layer with one output node with a linear activation function. The network used for the second dataset had a total of 5 dense hidden layers in place of the three used earlier. [5]

D. Convolution Neural Network

A convolution neural network (CNN) is a form of the artificial neural network specifically designed to process pixel data for image recognition and processing. In image processing, the Convolution Neural Network is Robust. Furthermore, CNN is also Artificial Intelligence, which uses Deep Learning to perform both productive and descriptive tasks. [4] In this analysis, we had to build a convolution neural network that regresses and gives a linear output that we can use as the problem we are solving is not a classification problem but rather a regression model. [2]b1

In our CNN-Regression model, we have one input layer which accepts an image of 400x400 size, one convolution 2d layers with a 3x3 filter and a max-pooling 2d layer with stride 2x2, two convolution 2d layers with a 3x3 filter, and a max-pooling 2d layer with stride 2x2, two convolution 2d layers with a 3x3 filter and a max-pooling 2d layer with stride 2x2, a flatten layer and three fully connected dense layers and finally an output layer which has one output with linear activation function [3].

III. MODELING AND ANALYSIS

The performance was analyzed based on various factors such as:

• Mean Absolute Error: Absolute Error is the amount of error in our measurements. It is the difference between the measured value and the "true" value. The goal is to minimize the MSE.

• Mean Squared Error: mean squared error (MSE) measures the average of the squares of the errors, that is, the standard of the squared difference between the estimated values and the actual value. MSE is a risk function corresponding to the expected value of the squared error loss. The goal is to minimize the MSE.

• R2-Score: It is (total variance explained by the model) / total variance. So if it is 100, the two variables are perfectly correlated, i.e., with no variance at all. A low value would show a low level of correlation, meaning a regression model that is not valid, but not in all cases. The lower the R2 score, the more accurate the model will be, But it is not precise in all cases.

IV. RESULTS AND DISCUSSION

All the parameters given in III-E are calculated for the comparison between different models. All the results presented below are calculated for the second dataset. In the case of Random Forest regression, the results are as follows:

• Mean Squared Error: 64.986
• Mean Absolute Error: 6.160
• R2 Score:-0.355

In the case of Support Vector regression, the results are as follows:

• Mean Squared Error: 76.452
• Mean Absolute Error: 6.869
• R2 Score:-0.912

In the case of Deep Neural Network, the results are as follows:

• Mean Absolute Error: 2.162
• loss: 11.148
• Validation Absolute Error: 2.049
• Validation loss: 10.961

In the case of the image dataset, CNN has applied the results are:

• Mean Absolute Error: 9.974
• loss: 145.040
• Validation Absolute Error: 13.771
• Validation loss: 221.351 All the given results can be visualized in section III-H

V. FIGURES

Neural Network: We can see from the given figures as the epochs in the case of Neural Networks increases the error decreases and in the end, we are achieving the model with the least MAE and max Accuracy.

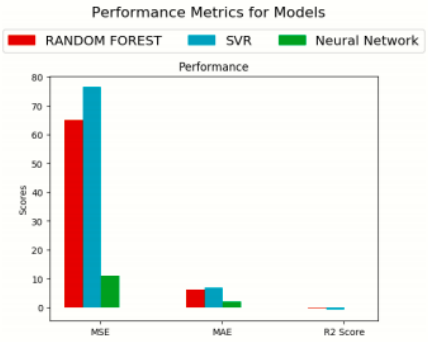

Fig. 1. MAE vs Epoch: CNN Regression

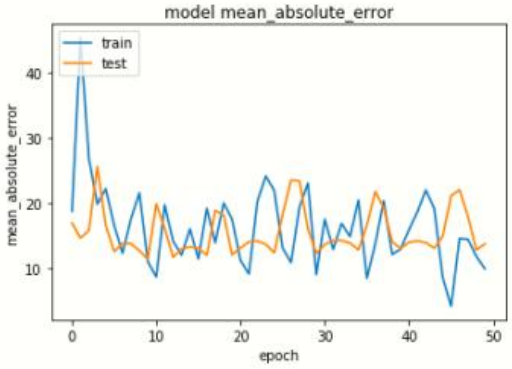

Fig. 2. MAE vs Epoch: CNN Regression

VI. CONCLUSION

From the results given in section III-F, we can conclude that Deep Neural Networks is preferably the best model for predicting the crop yield as it has the lowest Mean Absolute error. The same can be visualized from the graphs given in section III-H

In counting the crops from the images, the Accuracy is not very good due to the unavailability of sufficient data. We had only ten ideas, but the Accuracy can be improved, and the model can be made more accurate. The graph relating the MAE and epochs can be seen in section III-H.

REFERENCES

[1]. Crop Yield Prediction through different Machine Learning

[2]. Yuan, Jun Ni, Bingbing Kassim, Ashraf. (2014). Half-CNN: A General Framework for WholeImage Regression.

[3]. Keras, Regression, and CNN's, Adrian Rosebrock, January 28, 2019, unpublished.

[4]. Choudhury, A. Jones, J.. (2014). Crop yield prediction using time series models. Journal of Economics and Economic Education Research. 15. 53-68.

[5]. Alaslani, Maram Elrefaei, Lamiaa. (2019). Transfer Learning with Convolutional Neural Networks for IRIS Recognition. International Journal of Artificial Intelligence Applications. 10. 49-66. 10.5121/ijaia.2019.10505.

[6]. Ms. Kavita Joshi, Dr. D.D. Shah," Hybrid of the Fuzzy C Means and the Thresholding Method Segment to the Image in Identification of Cotton Bug, Volume 13, Number 10 (2018) pp. 7466-7471

[7]. Amuta Aware, kavita Joshi,"Wavelet Based Crop Detection And Automatic Spraying of Herbicides"International Journal of Innovations & Advancement in Computer Science, Volume 4, Issue 2,February 2015

[8]. Anish Polke, Kavita Joshi" Leaf Disease Detection based on machine learning, "International Conference on ISMAC in Computational Vision and Bio-Engineering (ISMAC – CVB 2018) and Springer - Lecture Notes in Computational Vision and Biomechanics.," May 16-17, (2018).

[9]. Joshi K., Shah D.D., Deshpande A.A. (2020), "Improvement in Satellite Images by Amalgam of Broveyand PCA Algorithm with Artificial Neural Network". In: Kumar A., Mozar S. (eds) ICCCE 2019.Lecture Notes in Electrical Engineering, vol 570. Springer, Singapore

[10]. M. K. V. Joshi, D. D. Shah and A. Deshpande, "Application of Fusion Technique and Support Vector Machine for Identifying Specific Vegetation Type", 2019 IEEE 5th International Conference forConvergence in Technology (I2CT), Bombay, India, 2019, pp. 1-5.

[11]. M. K. V. Joshi, D. D. Shah and A. Deshpande," Improving satellite image processing via hybridization of fusion, feature extraction & neural nets", International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Volume-7, Issue-6, March 2019 [6]

[12]. Amuta Aware, Kavita Joshi, "Crop and Weed Detection Based on Texture and Size Features and Automatic Spraying of Herbicides",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, January 201