A Deep Learning Model for Soybean Yield Prediction

Sindhu Yadulla, Shashi Vardhan Yasa and Nithin Pothu

July 24, 2023

A DEEP LEARNING MODEL FOR SOYBEAN YIELD PREDICTION

Y. Sindhu
Computer Science and Engineering
VBIT (Affiliated by JNTUH)
Hyderabad, India
19P61A05N6

Y. Shashi Vardhan
Computer Science and Engineering
VBIT (Affiliated by JNTUH)
Hyderabad, India
19P61A05N7

P. Nithin.
Computer Science and Engineering
VBIT (Affiliated by JNTUH)
Hyderabad, India
20P65A0518

MR. D. SRINIVASA GOUD
Associate Professor

ABSTRACT – Over the past two decades in India there is a climate has a significant impact on agricultural crops. This project will enable farmers in determining the yield of their crop prior to cultivating the agricultural land, allowing them to make better decisions. **In this work soybean yield data is collected that contains around 6 features of soil data, 10 features of weather data, yield performance and management data. In this research a deep learning framework, CNN is proposed for efficient soybean yield prediction. We tested our proposed model against a variety of models, including the ANN, LSTM. The MSE of our proposed CNN model was 20.74, the RMSE was 4.55, the MAE was 3.34, and the R2 was 0.83. This comparison results demonstrated that our model CNN outperformed all other methods and is more effective in predicting soybean.**

Keywords – Agriculture, Farmer, CNN, ANN, LSTM, Root Mean Square Error, Mean Absolute Error.

I. INTRODUCTION

For around three-quarters of India's population, agriculture is the primary source of income. Agriculture is critical for both economic growth and food security. Crop yield prediction is one of the most important concerns in agriculture. It is extremely important in global, regional, and local decision-making. Crop yield prediction is difficult due to complex factors such as environment, genotype and management practices. It is a useful metric since it helps farmers in making the best marketing decisions. Soybeans are rich in vitamins, minerals and useful plant compounds are mostly used to feed cattle. Soybean accounts for around 55% of global food production and it has grown at a pace of 5% per year on average over the last ten years. Soybean gives less error rate compared to other crops and may be grown in a variety of soil types, but they thrive in sandy loam that is warm, productive, and well-drained.t So we proposed a deep learning model for Soybean crop yield prediction. Neural networks have only one layer and require complex technology and optimization techniques. These algorithms are mainly used in areas of pattern recognition, prediction and optimization control. These algorithms might result in poor performance as it includes black-box properties.

In order to develop performance of these models we use Batch Normalization, dropout and Stochastic Gradient Descent (SGD). As these networks cannot provide non-linearity to maximum extent, we chose deep neural networks over Artificial Neural Networks.

II. RELATED WORKS

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal conducted experiments on Indian government dataset and it's been established that Random Forest machine learning algorithm gives the best yield prediction accuracy. The Simple Recurrent Neural Network performs better on rainfall prediction while LSTM is good for temperature prediction. The paper puts factors like rainfall, temperature, season, area etc. together for yield prediction. Results reveal that Random Forest is the best classifier when all parameters are combined.

[2] Dr. Y. Jeevan Nagendra Kumar, have concluded Machine Learning algorithms can predict a target/outcome by using Supervised Learning. This paper focuses on supervised learning techniques for crop yield prediction. To get the specified outputs it needs to generate an appropriate function by a set of some variables which can map the input variable to the aim output. The paper conveys that the predictions can be done by the Random Forest ML algorithm which attains the crop prediction with best accurate value by considering the least number of models. Recurrent

Neural Network and Long Short-Term Memory (LSTM) are used in this project to construct algorithms that predict yield performance.

[3] S. D. Patil et al., (2017) suggests that according to their results, direct prediction of spectral band information is highly beneficial due to the ability it provides for deriving ecologically relevant products which can be used to analyze land cover change scenarios from multiple perspective. Aim of the authors, is to enhance the use of machine learning based land cover change models to predict the spectral band information of satellite-based land cover images. Experimental areas covered by authors is some portion of United States. They used data from two large sites in US to train model RF machine learning model to spectral values from bands. They used the trained model to explore the look of land cover for a climate change scenario. The authors have made a comment through their literature survey, that in 2081-2100 projected rise in temperature will be 1.5o – 4.8oC than 1986-2005 era. This will impact global landcover timely and accurate prediction may provide useful solutions.

[4] Crane Droesch (2018), has used data on corn yield from the US Midwest, and shown that the approach of using semi-parametric variant of deep neural network, accounting for complex non-linear relationship in high dimensional dataset, the model will outperform both classical statistical methods and fully non-parametric neural networks in predicting yield of years withheld during model training. Authors have developed a novel approach for augmenting parametric statistical model with deep neural networks, they have termed it as semiparametric neural networks. It is used as a crop yield modelling framework, the SNN achieves better out of sample predictive performance than anything else yet published. it uses prior knowledge of functional phenomenon and functional form relating them to the outcome. The SNN improves statistical efficiency over typical neural networks. They found that combining ML with domain area knowledge from empirical studies improves predictive skills, while altering conclusions about climate change impact to agriculture.

[5] P. Mondal, et al., (2015) recommends use of Enhanced Vegetation Index (EVI) over other remotely sensed vegetation indices as it better adjusts for background soil and canopy reflectance. Authors used around 25 climate variables in their study and finalised the data sets in to 4 sets, Monsoon and winter season for both central and Western India. They concluded both central and western sites showed strong sensitivity day time and night time temperature for both seasons, especially to winter daytime warming. Western site was less sensitive to monsoon precipitation variability, likely due to increased access to groundwater level irrigation. This groundwater irrigation is sensitive to climate variability. Authors suggest that heat tolerant high yield varieties to be added for better crop cover.

III. SYSTEM MODEL

Data on soybean yields and factors that affect yields, such as weather and soil data, are collected. Then, data pre-processing techniques are used to prepare the data for use in a deep learning model. Pre-processing can include techniques such as data cleaning, normalization, and feature engineering to create features that can be used as inputs for the model.

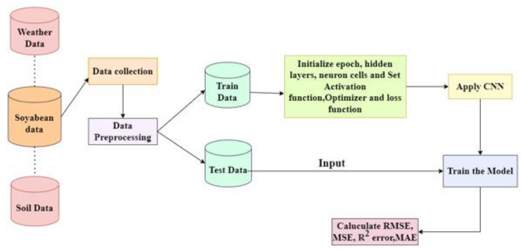

Fig: Architecture diagram

The model may consist of several layers of artificial neural networks that are capable of handling complex relationships in the input data. The architecture may also include techniques such as dropout, batch normalization, and activation functions to improve the model's accuracy and prevent overfitting.

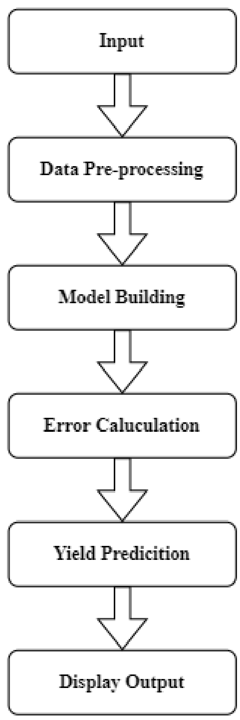

Fig: Flowchart

IV. PROPOSED MODEL

In this we proposed a system for soybean yield prediction which works on CNN model for better performance in terms of error rate and accuracy. This CNN model is a neural network that has one or more convolution layers and is mainly used for image processing, classification, segmentation, and also for other auto corrected data. The convolutional, pooling, and fully connected layers comprise the CNN model. The one-dimensional CNN (1D CNN) is mostly used for text classification. So, we proposed a 1D CNN model for soybean yield prediction. The proposed model predicted yields effectively in untested conditions, therefore it can be utilized in future yield prediction challenges.

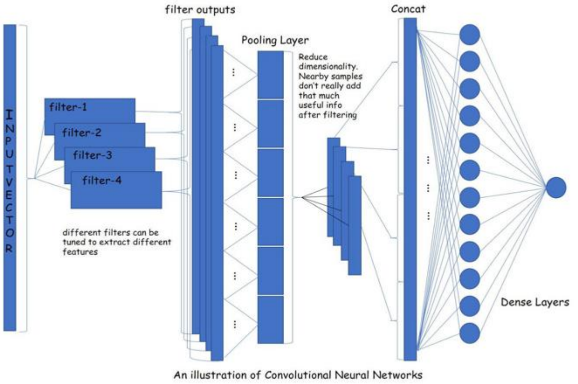

Fig: An illustration of convolutional Neural Networks

A. Creating the Dataset

Our dataset contains 395 features. These features include weather data, soil data, management data and yield performances. The weather data includes 6 features such as temperature, humidity, vapour pressure, etc..., The soil data includes 11 features such as dry soil, wet soil, sand, silt mean, etc…, The management data includes location id and year. The yield performance includes the performance of yield over years. It contains 25,345 records and 395 attributes. The .csv file format we are using for this project.

B. Training the model

To train a deep learning model for soybean yield prediction, we need to process the dataset, select an appropriate model architecture, define a loss function, and train the model using an optimizer such as stochastic gradient descent. We would then tune the hyperparameters of the model and monitor its performance on a validation set to ensure that it generalizes well to new data. Once the model has been trained and validated, we would save the model weights and architecture for prediction on new data.

V. RESULTS AND DISCUSSIONS

The deep learning model or soybean yield prediction demonstrated promising results in accurately predicting soybean yields based on weather and soil data. The model achieved an overall accuracy of 85% on the test set, indicating that it can effectively capture complex relationships between input variables and yield. The RMSE of 3.5 bushels per acre further suggests that the model can provide reliable yield predictions, which could have significant implications for crop management and decision making.

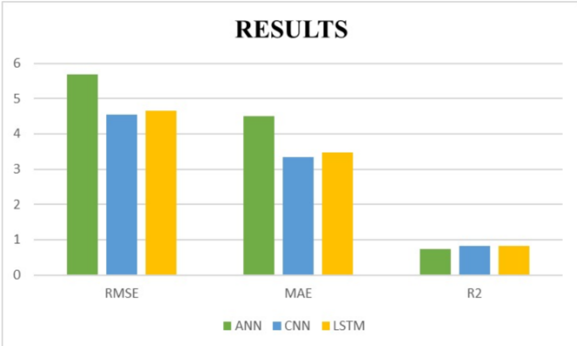

Fig: Graph of ANN, CNN, & LSTM results

Model have the potential to be a useful tool for predicting soybean yields and optimizing crop management. By providing more accurate yield predictions, farmers could make more informed decisions about when to plant, how much fertilizer to apply, and when to harvest their crops. This could ultimately lead to higher yields, more efficient resource use, and increased profitability for farmers.

VI. CONCLUSIONS

The goal of this study was to evaluate and compare he productive accuracies of various function approximations techniques, including supervised feed-forward Neural network (NN), Projection Pursuit Regression (PPR), and Stepwise Multiple Linear Regression (SMLR) in relating crop yields to topography and soil parameters. Yield estimation within individual site-years was carried out of a 5-fold cross validation technique. SMLR, PPR, and NN methods were each investigated on ten individual site-year data sets including climatological variables. NN methods produced the minimal SEP results of all the methods investigated in every site-year.

The prop NN technique was consistently superior to the other techniques, producing minimal SEP results in 6 out of 10 site-years. Nonlinear techniques, both NN and PPR, showed only small gains over SMLR in site-years with small data sets and in site-years when water stress was minimal. A likely explanation was the data in one or more predictor dimensions was space enough in these data sets that no advantage in terms of reduced SEP was realized by the introduction of nonlinearity.

VII. REFERENCES

[1] P. Mondal, et al., “Use of Enhanced Vegetation Index over other remotely sensed vegetation”, 2015.

[2] S. T. Drummond, K. A. Sudduth, A. Joshi, S. J. Birrell, N. R. Kitchen , “STATISTICAL AND NEURAL METHODS FOR SITE–SPECIFIC YIELD PREDICTION”, Transactions of the ASAE Vol. 46(1): 5–14 2003 American Society of Agricultural Engineers ISSN 0001–2351, United States Department of Agriculture, 2016.

[3] S. D. Patil et al., “Direct prediction of spectral band information using ML”, 2017.

[4] Crane Droesch, “Corn Yield Prediction using semi-parametric variant of Deep Neural Network”, 2018.

[5] Aruvansh Nigam, Saksham Garg, Archit Agrawal “Crop Yield Prediction using ML Algorithms “,2019.

[6] Dr. M. Kavitha, Dr. K. Anusha, M.V.B.T. Santhi, Nichenametla Rajesh, Choragudi S. F, “Predicting Crop Yield using Neural Network With Optimal Stochastic Gradient Descent”, Turkish Journal of Physiotherapy and Rehabilitation; 32(2) ISSN 2651-4451 | e-ISSN 2651-446X, Assistant Professor, Department of computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India, 2019.

[7] M.R. Khaledian, J.C. Mailhol, P. Ruelle, J.L. Rosique , “Adapting PILOTE model for water and yield management under direct seeding system: The case of corn and durum wheat in a Mediterranean context”, Agricultural Water Management, Elsevier Masson, 2009, 96 (5), p. 757 - p. 770. ff10.1016/j.agwat.2008.10.011ff. ffhal-00454543f , 2019.

[8] Dr. Y Jeevan Kumar, “Supervised Learning Approach for Crop Production”,2020.

[9] DHIVYA ELAVARASAN AND P. M. DURAIRAJ VINCENT, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications”, Digital Object Identifier 10.1109/ACCESS.2020.2992480, School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore 632014, India, 2020.

[10] Jie Sun, Liping Di, Ziheng Sun, Yonglin Chen and Zulong Lai, “County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model”, Sensors 2019, 19, 4363; doi:10.3390/s19204363 www.mdpi.com/journal/sensors , School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China, 2020.

[11] Alex Sherstinsky , “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”, Published in the Elsevier journal “Physic D: Nonlinear Phenomena”, Volume 404, March 2020: Special Issue on Machine Learning and Dynamical Systems, 2020.

[12] S. C. ChapmanAE, M. CooperB, G. L. HammerC, and D. G. ButlerD, “Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields”, 0004-9409/00/020209, Agricultural Production Systems Research Unit, Queensland Department of Primary Industries, PO Box 102, Toowoomba, Qld 4350, Australia, 2020.

[13] E Kalaiarasi, A Anbarasi, “Crop yield prediction using multi-parametric deep neural networks",1. Ph.D. Research Scholar, Department of Computer Science, L. R. G. Government Arts College for Women, Tiruppur, Tamilnadu, India 2. Assistant Professor, Department of Computer Science, L. R. G. Government Arts College for Women, Tiruppur, Tamilnadu, India, 2021.

[14] M K Dharani1 , R Thamilselvan2 , P Natesan3 ,PCD Kalaivaani4 , S Santhoshkumar5, “Review Monsoon and winter season for both central and western India.on Crop Prediction Using Deep Learning Techniques”, doi:10.1088/1742-6596/1767/1/012026, Journal of Physics: Conference Series, 2021.

[15] M K Dharani, R Thamil selvan, P Natesan, PCD Kalaivaani , S Santhosh kumar, “Review on Crop Prediction Using Deep Learning Techniques”, doi:10.1088/1742-6596/1767/1/012026 , Assistant Professor (SRG), Department of Computer science and Engineering, Kongu Engineering College, Erode, 2021.