International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 10, October 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Machine Learning Techniques on Crop Yield Prediction - Sustainable Agrarian Applications

[1]B Muralikrishna, [2]Dr P Latchoumi, [3]I. Sapthami

[1]PhD scholar, Dept of CSE, CRESCENT University, Chennai, Assistant Professor, Dept of CSE,PBR VITS, Kavali Mail:mkbvts@mail.com Ph:9989861882

[2]Associate Professor, Dept of IT, CRESCENT university, Chennai

[3]Research Scholar,Dept of CSE, SRM, Chennai

Abstract

In-season crop yield estimation has variousapplications such as the farmer taking corrective measures to increase the yield. We are exploring the efficient prediction of environmental parameters with help of machine learning and data mining techniques . The various data mining techniques are used on data for prediction of environment. The data is related to humidity, PH. value, water, soil type and atmospheric pressure these are responsible for crop yield. The efficient prediction is obtained by reading the environmental parameters and processed these through the machine learning algorithm,the processed environmental parameter values are useful for farmers to take decisions about further implantation of crop yield. A specific crop is also predicted with for particular region and season also. One of the most important fields is decision tree. By analyzing the soil, water levels, PH values and atmosphere at particular region best crop is Predicted.This prediction will help the farmers to choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months. This prediction can be carried out using Random Forest classification machine learning algorithm.

Keywords: Crop yield, Fertilizers, Humidity, Machine Learning Techniques, PH,Soil.

I INTRODUCTION

Crop yield expectation is a significant agrarian issue. The Agricultural yield principally relies upon climate conditions, pesticides. Exact data about history of harvest yield is significant for settling on choices identified with farming danger the board and future expectations. The study of preparing machines to learn and create models for future forecasts is generally utilized, and not to end. Agribusiness assumes a basic job in the worldwide economy.. With the effect of environmental change in India, majority share of the agrarian yields is by and large severely influenced regarding their presentation over a time of most recent two decades. Anticipating the harvest yield well in front of its gather would help the measures for promoting and capacity. Such expectations will likewise help the related arrangement producersand ranchers for taking fitting. Enterprises for arranging the coordination's of their business. Harvest creation is a mind-boggling wonder that is impacted by climatically input parameters. Agribusiness input parameters differs from field to field and rancher to rancher. Gathering such data on a bigger region is an overwhelming errand. Nonetheless, the climatic

data gathered in India at each square meter territory in various pieces of the zone arranged by Indian Meteorological Department. Additionally, the yield of each harvest in each stateisgatheredanddistributedbythebranchofagribusiness and collaboration consistently. Such informational collections are utilized right now foreseeing the impact on significant harvests and along these lines, their yield in a future year.

II RELATEDWORK

Anil SuatTerliksiz et.al., concentrated on soybean yield forecast of Lauderdale County, Alabama, USA utilizing 3D CNN model that use the spatiotemporal highlights [1]. (it is [2]) The yield is given from USDA NASS Quick Stat apparatus for a considerable length of time 2003-2016. The expectation of harvest yield has direct effect on national and worldwide economies and assume significant job in the nourishment the executives and nourishment security.

Niketa Gandhi et.al. [2] Proposed a choice emotionally supportive network model for rice crop yield forecast for Maharashtra state, India. A GUI has been made in Java utilizing NetBeans apparatus and Microsoft Office Access database for the simplicity of ranchers and leaders. The interface takes into account the determination of the scope of precipitation, least temperature, normal temperature, most extreme temperature and reference crop evapotranspiration and predicts the normal class of yield viz., low, moderate or high. Ranjini B Guruprasad et.al.,[3] introduced a contextual analysis of climate and soil information-based yield estimation demonstrating for paddy crop at various spatial goals(SR)levels, to be specific, at the area and taluk levels in India. We give a point by point investigation of precision of the yield estimation models across changed arrangements of highlights and diverse AI systems. Nilima et.al., [4] introduced a thought for example to how to send WSN on fieldandhowMachinelearningmodelisfittedforforecastof bug/ailments utilizing Naive Bayes KernelAlgorithm.

Predicting Crop yield and Effective use of Fertilizers using Machine Learning Techniques

Remote Sensor Network is new innovation to world and nation like India where it can utilize in Agriculture Sector in India for expanding yield by giving early expectation of plant sicknesses and bug. This can be occurred by taking crude information from field where WSN organize is introduceand with fitting proper AI model for this information to get anticipatedyield.

Shruti Kulkarni et.al., presents a model for example an information driven model that learns by notable soil just as precipitation information to break down and anticipate crop yield over seasons in a few locales, has been created [5]. For this investigation, a specific yield, Rice is considered. The planned half breed neural system model distinguishes ideal mixes of soil parameters and mixes it with the precipitation design in a chose locale to develop the expectable harvest yield. The spine for the prescient investigation model regarding the precipitation depends on the Time-Series approach in Supervised Learning.

S. Bhanumathi et.al. Analyses the different related characteristics like area, pH esteem from which alkalinity of the dirt is resolved. Yield forecast is a significant issue in rural.Anyrancheriskeenonknowinghowalotofyieldheis going to expect [6]. Every one of these characteristics of information will be dissected, train the information with different appropriate AI calculations for making a model. Neha Rale et.al. [7] Propose to utilize AI

procedures to build up an expectation model for crop yield creation. They analyses the exhibition of different direct and non-straight regressor models utilizing 5-overlap cross approval. Previously, ranchers used to foresee crops dependent on their own understanding and watched climate conditions.Climate, irritations, and collect activity might be kept as reference for futureyears.

T.Mhudchuayet.al.[8] Concentrated on down pour tookcare of rice where the fundamental activities are when to begin development and when to collect. The objective is to locate the ideal development and collect period to such an extent that ranchers' salary is amplified. This paper speaks to a use of a Deep Q-learning in the rice crop development practice, where the ideal activities are resolved.

Shivi Sharma et.al., [9] proposed a technique utilized, in that dirt and condition highlights for example normal temperature, normal stickiness, all out precipitation and creation yield are utilized in anticipating two classes in particular: great yield and awfulyield.

Suhas S Athani et.al. [10] Presents the data relating to the harm of harvests as of late because of the development of weeds. Weeds are one of the significant hazards to the genuine home and mankind. Right now, thought, Support Vector Machine (SVM) Classifier is used to make out whether plant is harvest or weed. The maize crops are consistently observed by catching pictures utilizing camera. So as to group a plant as a yield or weed, different highlights are removed which among them are shape, surface, shading.

III METHODOLOGY

Predicting crop yield using the powerful algorithm and determining how much fertilizer should be used to get the crop's proper yield.

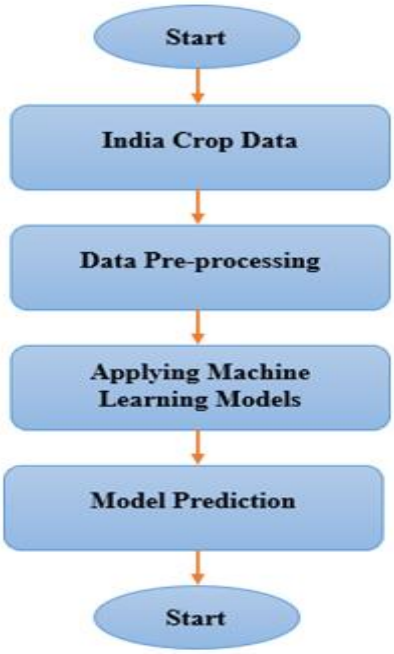

Fig 1. Process of Our ProposedMethodology

In methodology consists of following phases:

- Overview ofData
- DataPreprocessing

- ModelSelection
- CropPrediction
- Required Packages andLibraries

A. Overview of Data

In this paper, India crop data set was used i.e. is for prediction. This is the collection of sample data used in this project. The data used to estimate yields of crops based on 9 variables. We can construct a machine learning model and trainthemodelusingthese9factorsandwepredicttheoutput and we can predict from the data set how much fertilizer will be used to achieve the correct yield.

B. DATASET AND FEATURES

To perform the crop yield prediction task with remotely senseddata, we leveraged Moderate Resolution Imaging Spectroradiometer(MODIS) satellite imagery, which provides free and easy to accesscoverage of the entire globe. Of the many imagery products exportedby MODIS, we used MOD09A1 [10], MYD11A2 [11], andMCD12Q1 [1], which provide, respectively, eight-day compositesfor seven-band reflectance imagery, two-band daytime and nighttimetemperature imagery, and a land cover mask.

Each reflectance bandrepresents a distinct range of velengths sensed by the MODISsatellite. The land cover mask is updated annually and was only usedto distinguish cropland from non-cropland. As our ground truth for soybean crop yields, we used county level and province-level yield statistics compiled by the ArgentineUndersecretary of Agriculture [9] and the Brazilian Institute ofGeography and Statistics [8]. All yields were reported in units ofmetric tonnes per cultivated hectare (t/Ha).Each officially reported crop yield of a particular region for aparticular harvest was paired with a sequence of MODIS reflectanceand temperature images from the months preceding the harvest. Inorder to train our deep learning models, we processed the MODISimagery into the dimensionally reduced pixel histograms describedby You et al. For each image $I \in Rh \times w \times d$, where h, w, and d arerespectively the number of image height pixels, width pixels, andreflectance/temperature bands, we assumed each band to be independent
of all others and created d pixel histograms. For each histogram,we generated a number of buckets that represented different rangesof reflectance values or temperatures and placed pixels into theircorresponding buckets. We grouped these individual histogramstogether, so the final representation of I was a two-dimensional matrix$H \in Rb \times d$, where b is the number of bins we chose. For eachsoybean harvest, we stacked histogram matrices from the imagesequence associated with that harvest into a single three-dimensionaltensor.

Because the MODIS cropland mask does not distinguish soybeansfrom other crops, we ignored regions that contributed the bottom5% of total production in Argentina and Brazil in order to onlytrain our models on regions with significant soybean crop cover andalso filter out noisy crop yield values from regions with very lowsoybean production. Unfortunately, this eliminated a large number of harvests from our datasets, increasing the difficulty of the yieldprediction task, especially in Brazil. Our Argentine dataset contained1,837 harvests after filtering, and our Brazilian dataset contained336 harvests after filtering.

C MODELS

Baseline Models

Our baseline models used ridge regression with varying regularizationconstants. Using histograms directly for ridge regression was infeasible due to their high dimensionality. We leveraged two differentmethods of feature extraction to reduce dimensionality. Inthe "band mode" method, we used only the mode of each band'shistogram slice at each time step, creating an input vector of length dt for a single harvest. As an example, the first band of the histogramtime series shown in Figure 1 would be replaced by a vectorcontaining the bin index that is the brightest at each time step. Ourother feature extraction method utilized the Normalized Difference Vegetation Index (NDVI) due to the metric's prevalence in industry.For each input sequence, we calculated the mean NDVI at each timestep, yielding a feature vector of length t .

Deep Learning Models

Our primary model was a recurrent neural network composed oflong short-term memory (LSTM) cells. We flattened the three dimensional pixel histogram representation of a region into a two dimensional matrix by concatenating on the MODIS bands andhistogram bucket dimensions, which preserved the time dimension. The LSTM layer took d histograms as input at each time step andsent its ultimate activation to a final dense layer, which output thepredicted crop yield. A visualization of our model architecture isshown in Figure 1.For transfer learning from Argentina to Brazil, we initialized theLSTM model with the parameters from a neural network trained on Argentine soybean harvests. We stripped out the last dense layerof the pre-trained model and replaced it with an untrained denselayer of the same dimensions before training the modified modelon the available Brazilian training data. In this manner, we fully recalibratedthe last dense layer and fine-tuned the rest of the Argentinemodel's parameters.

D RESULTS AND DISCUSSION

We trained and tested our models on Argentine soybean harvestsfrom 2012 to 2016 in order to evaluate the efficacy of the pixelDeep Transfer Learning for Crop Yield Prediction with Remote Sensing Data COMPASS '18, June 20–22, 2018 histogram and LSTM model approach in a developing country withless available data. For each testing year, we trained the modelon harvests from all years except for that year. Learning rates andstopping criteria were tuned on a hold-out validation set sampledfrom 20% of the training data. A comparison of deep learning andbaseline RMSE values can be found in Table 1 and the associated R2values for the LSTM models are in Table 3. On average, the LSTM models outperformed ridge regression aselines, demonstrating the utility of the approach. Notably, we observed that 2014 is an outlier year with a negativeR2 score for both the neural network model and the baselines. NegativeR2 scores occur when a model does not follow the trend of thedata. This performance was likely due to the fact that the test setwas not sampled randomly from the full population of harvests fromall years but was instead sampled from a single year, in this case2014. Any anomalies localized to that year, such as unusual weatherpatterns like the onset of the strong 2014-2016 El Nino event [5]or social factors, could have disproportionately impacted soybeanyields and led to poor generalization to this specific test year. In addition, we trained our model to forecast Argentine soybean crop yields in advance of the harvest date. For example, in order toforecast the soybean yield four months (about 50% of the season) in advance of the harvest in June, we withheld the second halfof the image sequence

corresponding to that harvest, training andpredicting only using the first half. Figure 2 shows the results of thisforecasting strategy with season fractions ranging from 25% to thefull 100%.

Performance on the test year was in general best with the full data available during training, but even with access to only 25%of the information, our model exhibited positive predictive power inmany years.As expected, we saw that the predictive performance of our modelon the training and development sets increased monotonically as weprovided more data. However, this was not true of the test set in all cases. Namely, there were sometimes dips in R2 (or, equivalently, jumps in RMSE) when predicting with 50% and 75% of the data. Once again, this was likely due to the nonrandom nature of the test set. It is possible that providing additional mid-season data helpedthe model learn features that were important to the training and developmentsets but were irrelevant to the given test set. Interestingly,performance sometimes improved when 100% of the season datawas provided even after experiencing dips with 50% and/or 75%This suggests that the beginning and end conditions of a harvest arethe most consistent indicators of yield between years.

Brazil

Given the promising performance in Argentina, we turned our attentionto Brazilian soybean harvests to test our model's transferabilityto different regions. As reference points, ridge regression baselinesand an LSTM model were trained on only Brazilian soybean harvestsfrom 2012 to 2016 while a separate LSTM model used transferlearning from Argentina.

Our transfer learning model outperformed all other models thatwere trained only on Brazilian data. A comparison of RMSE valuesare shown in Table 2. A more detailed breakdown of RMSE andR2 scores for our standard and transfer learning models is shown inTable 3.Figure 2: Argentine soybean harvest forecast performance on test years 2012-2016 as a function of season data HarvestLSTM;histogramsRegression;NDVI Regression; band modes

2012 0.54 0.60 0.64

2013 0.60 0.59 0.67

2014 0.73 0.75 0.75

2015 0.54 0.94 0.93

2016 0.70 0.92 1.04

Table 1: RMSE of LSTM and baseline models

ExperimentsHarvest

LSTM; histograms LSTM; histograms & transfer learn Regression; NDVI Regression;band modes

2012 0.42 0.38 0.56 0.68

2013 0.29 0.26 0.40 0.60

2014 0.23 0.26 0.28 0.33

2015 0.53 0.50 0.54 0.60

2016 0.62 0.52 0.49 0.73

Table 2: RMSE of LSTM and baseline models

CONCLUSIONS AND FUTURE WORK

This paper presents a preliminary deep transfer learning frameworkfor reliable crop yield prediction in developing countries with remotesensing data. The results in Argentina and Brazil demonstrate thatthis approach can successfully learn effective features from rawdata and achieve improved performance compared to traditionalmethods. The ability to improve predictive performance in regionswith limited data by using transfer learning is exciting because

these regions especially stand to benefit from a cheap, reliable croppredictiontool. Next steps include expanding the application of this approach to new regions, supporting more crops, and using modelspre-trained on the United States, which has a significant amount ofreliable data, to transfer learn to other countries.

REFERRENCES

[1] Michael McPhaden and Aaron Levine. 2016. NOAA: How the failed 2014-15El Nino fueled the strong 2015-16 El Nino. https://www.pmel.noaa.gov/elnino/news-story/how-failed-2014-15-el-nino-fueled-strong-2015-16-el-nino

[2] Reid Pryzant, Stefano Ermon, and David Lobell. 2017. Monitoring EthiopianWheat Fungus with Satellite Imagery and Deep Feature Learning. In 2017 IEEEConference on Computer Vision and Pattern Recognition Workshops (CVPRW).IEEE, 1524–1532.

[3] N. A. Quarmby, M. Milnes, T. L. Hindle, and N. Silleos. 1993. The use ofmulti-temporal NDVI measurements from AVHRR data for crop yield estimationand prediction. International Journal of Remote Sensing 14, 2 (1993), 199–210.https://doi.org/10.1080/01431169308904332

[4]BrasilSistema IBGE de RecuperacaoAutomatica, InstitutoBrasileiro de Geografiae Estatistica. [n. d.]. Producao Agricola Municipal: producao das lavouras
temporarias. https://sidra.ibge.gov.br/tabela/1612

[5] Delegaciones y EstudiosEconomicos Argentina Subsecretaria de Agricultura, DireccionNacional de Estimaciones. [n. d.]. Datos Argentina: Estimacionesagricolas.
http://datos.gob.ar/dataset/estimaciones-agricolas

[5] Eric Vermote. 2015. MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3Global 500m SIN Grid V006. (2015). https://doi.org/10.5067/modis/mod09a1.006

[6] Zhengming Wan and S. Hook. 2015. MYD11A2 MODIS/Aqua Land SurfaceTemperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. (2015).
https://doi.org/10.5067/modis/myd11a2.006

[7] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. 2016.Transfer learning from deep features for remote sensing and poverty mapping. In
Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAIPress, 3929–3935.

[8] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. 2017.Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. 2017 Association for the Advancement of Articial Intelligence (2017).