IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

A Survey on Different Data Mining Techniques for Crop Yield Prediction

R. Beulah

[1]Department of Computer Science, TSAAST College, Coimbatore, Tamilnadu, India

[*]*Corresponding Author: compucare98@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 25/Jan/2019, Published:31/Jan/2019

Abstract— Crop growing is measured as the stamina of India, is the improvement of plant for foodstuff, bio-fuel, counteractive plants and other harvest for behind and enhancing human life. Farming is an unique business crop creation which is contingent on different attributes such as soil, climate, irrigation, precipitation, insect killer weeds, fertilizers, nurturing, temperature, harvesting and other factors. An accurate crop yield prediction helps support decision makers in the agriculture sector to envisage the yield effectively. Data mining techniques play a vital role in the study of data for crop yield prediction. Data mining is the computing method of discovering patterns in hefty datasets involving methods at the connection of machine learning, artificial intelligence, record and system database. This piece of writing presents a detailed examination of various techniques planned for crop yield prediction. At first, dissimilar techniques developed by previous researchers are calculated in detail. Then, a relative analysis is carried out to know the precincts of each technique and afford a suggestion for further enhancement in crop yield prediction successfully.

Keywords—Agriculture, crop yield prediction, data mining, machine learning technique**.**

I. INTRODUCTION

In India, Agriculture acts a strategic responsibility in the economy of the country. Due to copious reasons, the preponderance of farmers is not getting the probable crop yield. The crop yield depends on various factors such as soil, climate, irrigation, rainfall, pesticide weeds, fertilizers, cultivation, temperature, and harvesting. The farmer needs a sagacious guidance to envisage the future crop yield and a study is to be made to help farmers to exploit their crop production. India 10% to 30 % of the total vegetable yield is smashed yearly by diseases[13]. Yield prediction [1] is an important problem in the field of agriculture. Every farmer is involved in expressive about the crop yield. In the past, yield calculation was performed by analyzing farmer’s earlier practice on a meticulous crop manually. But, the volume of agriculture statistics is high and it is exceedingly composite to analysis physically.

Data mining techniques [2] are lengthily functional to the agricultural danger. It is used to examination a huge dataset and establishes serviceable classifications and patterns in the datasets. The overall reason of the data mining method is to extract useful information from the dataset and exchange it into an explicable structure for additional use.

The mined information is usually represented as a model of the semantic structure of a record, wherein the duplication may be used on new data for prediction or classification of farming data. The main purpose of this editorial is studying in detailed information on different techniques utilized for crop yield prediction. In addition, their limitations are addressed to further evolution the crop yield prediction successfully.

The rest of the editorial is deliberate as follows: Section II provides the earlier researches associated to crop yield prediction using diverse data mining techniques. Section III compares the recital competence of those crop yield prediction techniques and Section IV concludes the survey that reviews an entire argument.

II. SURVEY ON CROP YIELD PREDICTION TECHNIQUES

A Bayesian Network [3] was obtainable for prediction of rice crop yield in Maharashtra state India. Originally, 27 districts of Maharashtra were selected on account of available data from widely available Indian Government records with different climate and crop parameters. From the collected parameters, smallest amount temperature, reference crop desertion, highest temperature, average temperature, and area production were preferred for better crop yield prediction. The chosen parameters were given as contribution to BayesNet and NaiveBayes. These classifiers predicted the yield of yield effectively.

A parameter-based modified Artificial Neural Network (ANN) [4] replica was proposed for crop yield calculation. In the proposed model, there are a duo of methods were used to predict the yield of wheat. ANN and Multiple Linear

Regression (MLR) were used in this form. There are unusual parameters such as a quantity of precipitation, soil transpiration, crop biomass, measure of manure applied, soil desertion and Extractable Soil Water (ESW) were extracted and specified as input to ANN and MLR to predict the yield of wheat. Additionally, Default- ANN (D-ANN) and Customized- ANN (C-ANN) was used for wheat yield prediction. D-ANN is an ANN with only one hidden layer and C-ANN was developed by varying the amount of hidden layers, Learning Rate (LR) and a number of neurons in the hidden layer.

Machine learning and superior sensing techniques [5] were introduced for wheat yield prediction. There are three Self Organizing Map (SOM) based models called Counter-Propagation Artificial Neural Network (CP-ANN), Supervised Kohonen Networks (SKN) and XY-Fusion (XY-F) was used for wheat yield prediction. These three machine learning techniques used supervised learning to connect high-resolution data on crop and soil with isofrequency program of wheat yield productivity. Initially, physicochemical soil parameters were together with an online visible and near-infrared spectroscopy sensor, which was consequently incorporated with crop growth indicators considered with satellite imagery.

A hyperspectral calculation model of soil salinity using Partial Least Squares Regression (PLSR) [6] was obtainable for prediction of soil salinity in Tianjin coastal area. Soil salinization is single of the soil degradation processes which reduced the agricultural yield. An Analytical Spectral Device (ASD) Field Spec spectrometer was used to calculate the soil spectral reflectance of soil samples which are varied in salinity. The First-Order Derivative Reflectance (FDR) and treated Continuum-Removed (CR) were used and compared to discover the more constructive predicting model of soil salinity, which detected flimsy differences in spectral assimilation features compared with inventive reflectance. By using the PLSR model based on treated soil spectra, the first resultant CR reflectance was the most favorable spectra indexes and for this reason prediction exactness of the most advantageous PLSR model was enhanced.

A Decision Support System for Agro Transfer 4.6 (DSSAT) model [7] was introduced to replicate wheat yield and top soil organic carbon under a wheat-maize cropping system in the North China plain. In this model, the significance of low nitrogen input on wheat yield, grain nitrogen deliberation, and soil Organic carbon was investigated by using DSSAT beside with the CENTURT soil model. From the analysis, it was recognized that for N0 treatment the replication wheat yield and SOC were lesser than the calculated data. More than cultivar parameters soil parameters were more perceptive in crop growth.

A fuzzy logic and regression model [8] was introduced for crop yield forecasting. Time series data plays a significant role in the farming yield prediction. The fuzzy logic and regression model was alert on predicting data values on a large spectrum of fuzzy logic computations based on second and third-degree relationships. It worked on four special types of the fuzzy interval, where each interval was tested with four degrees of regression equations. Each of these sixteen cases was performed for the fuzzy logic relationship (FLR) two and three separately. This model forecasted the manufacture of wheat by using definite production as the universe of disclosure and intervals based partitioning.

A binary tree based machine learning method called Random Forest [9] was introduced to predict the crop give way based on climate and biophysical variables at global and local scales in potato, maize, and wheat in comparison with Multiple Linear Regression (MLR). Initially, data were composed from different sources are maize silage and potato tuber from the northeastern seaboard region, gridded global wheat grain yield and maize grain yield from US countries over thirty years. Then the collected data are trained by using Random Forest. A Random forest was used for both classification and regression. Here, it was used as a regression tool. Based on the trained data, the yield of maize, wheat, and potato was predicted.

A Convolutional Neural Network (CNN) [10] was introduced for bitter melon yield prediction. Based on descriptions of bitter melon leaf, the leaves were classified as high-quality and bad. CNN is a type of neural network which was more correct and it has the ability to train data which is plentiful in size. It enabled a network to increase its layer, learning more, learning in accuracy and decreasing in error. The computational process of CNN is described as finding the best set of weights for the neural network which was called as learning or training. According to the training data, the leaves were classified as good and bad and predict the yield of bitter melon.

A Support Vector Machine (SVM) with features role analysis [11] was introduced for undeveloped yield prediction. This predictive model of crop yield provided a structure for agricultural decision making and understanding how a variety of features affect yield. SVM was based on the perception of conclusion plane boundaries. A decision plane is one that separates between a set of objects having a different class membership. The main intention of this model was improving comprehensibility. But, SVM regression was a black-box model which has lack explanation of prediction. This problem was resolved by extending SVM with the newest algorithms to clarify regression models through analysis of features contributions.

A Bayesian model averaging (BMA) [12] was accessible for a numerous crop growth model to provide a more consistent prediction of maize yields. In order to separately generate unique predictions of country-level maize yields, the unusual models such as nitrogen oriented DeNitrification and DeComposition (DNDC), the photosynthesis oriented WOrld FOod STudy (WOFOST) model and the water oriented AquaCrop model was functional. Then, an addition calculation was achieved by using a linear arrangement of the three group members using BMA weights. The incorporated model was more accurate and precise predictions of maize yield than any individual model. The BMA model was remunerated the uncertainty of individual model effectively.

III. Results and Discussions

This segment presents an attribute about the merits and demerits of extraordinary crop yield calculation techniques whose realistic information is discussed in the preceding part. Through the appraisal of crop yield prediction by means of different data mining techniques, the next challenges are addressed. The Bayesian network model was predicted only rice yield at Maharashtra state. The MLR was not captured the non-linear association between the diverse input parameters. The most important drawback of SKN was it concerns the powerlessness to model permanent production relations. In addition, the PLSR is lack of representation test statistics. The N function in DSSAT-CENTURT would not be sustainable for crop yield prediction.

The fuzzy logic and regression model still needs an improvement in terms of means square incorrectness. The random forest based crop yield prediction has the risk of overfitting data for the conditions where training data were strenuous while its accuracy can diminish where training data were sparse. The CNN based crop yield calculation requires added features to improve prophecy accuracy. The Support Vector Machine is not much competent for crop yield prediction. In Bayesian model averaging, doubts consequential from elucidation and model input are not fully explored. From the following Table 1, the most exigent issues in crop yield calculation are experimental and a perfect resolution is ordinary to overcome those issues for crop yield prediction.

Table. 1 Comparison of Different Crop Yield Prediction Techniques

Ref. no.	Methods	Merits	Demerits	Performance Metrics
[3]	BayesNet, NaiveBayes	Bayesian Networks get better the decision support system for vital prediction of crop yield	It calculate only rice yield at Maharashtra state	BayesNet: Accurateness = 97.53 Sensitivity = 96.31% Specificity = 98.16% Mean Absolute Error (MAE) = 0.0425 Root Mean Squared Error (RMSE)= 0.1449 Relative Absolute Error (RAE)= 9.56% Root Relative Squared Error (RRSE)= 30.71% F1 Score = 0.96 Mathews Correlation Coefficient (MCC) = 0.94 NaiveBayes: Accuracy = 84.69 Sensitivity = 77.14% Specificity = 88.48% Mean Absolute Error (MAE) = 0.1456 Root Mean Squared Error (RMSE)= 0.2999

				Relative Absolute Error (RAE)= 32.74% Root Relative Squared Error (RRSE)= 63.58% F1 Score = 0.77 Mathews Correlation Coefficient (MCC) = 0.66
[4]	Artificial Neural Network, Default-Artificial Neural Network, Customized Artificial Neural Network, Multiple Linear Regression	Successfully predict wheat yield	MLR was not confine the non-linear relationship between the dissimilar input parameters	R^2: MLR = 92.52% D-ANN = 95% C-ANN = 97% Average Prediction Error: MLR = 4.196% D-ANN = 2.2408% C-ANN = 0.5275%
[5]	Supervised Kohonen Network, Counter-Propagation Artificial Neural Network, Supervised Kohonen Networks, XY-Fusion	Most excellent Accuracy	It fear the inability to model unremitting output relations	Prediction Accuracy SKN = 91.3% CP-ANN = 91.48% XY-F = 92.15%
[6]	Partial Least Squares Regression	Present a additional advanced hyperspectral predicting Model	PLSR is not have of model test statistics	Exactness = 94.4%
[7]	Decision Support System for Agro Transfer, CENTURT soil model	Helpful tool for measure and predicting wheat yield, grain N attentiveness, and SOC trends below wheat-maize cropping system	N application would not be sustainable	DSSAT model (N concentration): R^2 =0.62 Forecasting Efficiency = 0.81 normalized Root Mean Square Error = 16% DSSAT model (N150 treatment): R^2 =0.24 Forecasting Efficiency = -0.21 normalized Root Mean Square Error = 11%
[8]	Fuzzy logic and regression model	An inerrant and competent way to consider, evaluate and approximate wheat production	Still requirements development in conditions of mean square error	FLR second degree (Quartic): Mean Square Error (all degree in 5 intervals) = 189914 Mean Square Error (all degree in 7 intervals) = 189978 Mean Square Error (all degree in 9 intervals) = 224548 Mean Square Error

				(all degree in 11 intervals) = 222158
[9]	Random Forest	Easiness of utilize and efficacy in data analysis	RF has the danger of overfitting data for the situation where training data were determined while its accuracy can moderate where training data were thin	Root Mean Square Errors = 6-14%
[10]	Convolutional Neural Network	Requires a reduced amount of training time	Needs added features to pick up prediction exactness	Training time (50 iterations)= 74.24 seconds
[11]	Support Vector Machine	Simplify well level with restricted training samples	Fewer efficient	Maize dataset: Root Mean Square Error = 538.69 Mean Absolute Error = 443.17 Correlation Coefficient = 0.9220 Soybean dataset: Root Mean Square Error = 297.77 Mean Absolute Error = 228.49 Correlation Coefficient = 0.8337 Sugarbeet dataset: Root Mean Square Error = 4276.90 Mean Absolute Error = 3111.75 Correlation Coefficient = 0.8452
[12]	Bayesian model averaging	Reduce the uncertainty resultant from a assorted copy structure	Uncertainties resulting from clarification and model input are not fully explored	R^2: WOFOST = 0.74 DNDC = 0.63 AquaCrop = 0.71 BMA = 0.81 Root Mean Square Error (kg ha^1): WOFOST = 1000 DNDC = 1300 AquaCrop = 1110 BMA = 850 Absolute Error (kg ha^1): WOFOST = 780 DNDC = 1000 AquaCrop = 910 BMA = 730

IV. Conclusion

During this study an image denoising method based on clustering and the mishmash of PDE, FFT and dominates color has been planned. The exertion of every system is unusual so it is applied differently stage wise. It has been incorporated terms of clamor rate to determine the effectiveness in terms of PSNR values. The clatter fraction measured here is in the range of 1-20%. Hierarchical clustering for the intention based data parting has been applied. Then disintegration is performed with the comparison matching and place of the image element. The obtained PSNR standards put forward that our grades are better in assessment to the pervious approach.

References

[1] Bhanose, S. S., Bogawar, K. A., Dhotre, A. G., & Gaidhani, B. R. *"Crop and Yield Prediction Model",* International Journal of Advance Scientific Research and Engineering Trends, 1(1), pp.**23-28, 2016.**

[2] Ramesh, D., & Vardhan, B. V. *"Analysis of crop yield prediction using data mining techniques*", International Journal of Research in Engineering and Technology, 4(1), pp.**47-473**, **2015**

[3] Gandhi, N., Armstrong, L. J., & Petkar, O. *"Predicting Rice crop yield using Bayesian networks",* In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on pp. **795-799, 2016.**

[4] Shastry, K. A., Sanjay, H. A., & Deshmukh, A. *"A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction",* Journal of Artificial Intelligence, **vol,9**, pp. **23-32, 2016**.

[5] Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. *"Wheat yield prediction using machine learning and advanced sensing techniques",* Computers and Electronics in Agriculture, 121, pp.**57-65, 2016.**

[6] Wang, J., Li, Z., Qin, X., Yang, X., Gao, Z., & Qin, Q. (2014, July). *"Hyperspectral predicting model of soil salinity in Tianjin costal area using partial least square regression",* In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International pp. **3251-3254, 2014.**

[7] Liu, H. L., Liu, H. B., Lei, Q. L., Zhai, L. M., Wang, H. Y., Zhang, J. Z., ... & Liu, X. X. *"Using the DSSAT model to simulate wheat yield and soil organic carbon under a wheat-maize cropping system in the North China Plain"* Journal of Integrative Agriculture, 16(10), pp. **2300-2307, 2017.**

[8] Garg, B., Aggarwal, S., & Sokhal, J. *"Crop yield forecasting using fuzzy logic and regression model".* Computers & Electrical Engineering, **2017**.

[9] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. *"Random forests for global and regional crop yield predictions",* PLoS One, 11(6), e0156571. **2016.**

[10] Villanueva, M. B., & Salenga, M. L. M. *"Bitter Melon Crop Yield Prediction using Machine Learning Algorithm",* International Journal of Advanced Computer Science and Applications (IJACSA), 9 (3), pp.**1-6, 2018**

[11] Brdar, S., Culibrk, D., Marinkovic, B., Crnobarac, J., & Crnojevic, V. (2011, *"Support vector machines with features contribution analysis for agricultural yield prediction. In Second International Workshop on Sensing Technologies in Agriculture",* Forestry and Environment (EcoSense 2011), Belgrade, Serbia pp. pp.**43-47, 2011.**

[12] Huang, X., Huang, G., Yu, C., Ni, S., & Yu, L. *"A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging",* Field Crops Research, 211, pp.**114-124, 2017.**

[13] Asha Patil, Kalyani Patil, Kalpesh Lad, "*Leaf Disease detection using Image Processing Techniques*", International Journal of Scientific Research in Computer Science and Engineering" Vol:6 pp.**33-36, 2018.**

Authors Profile

Mrs. R Beulah pursed Bachelor of Science from Bharathiar University, Coimbatore in 2002 and Master of Science from Bharathiar University in year 2004. she is currently studying Ph.D. and she has published more than 3 research papers in reputed international journals including IEEE conferences and it's also available online. Her main research work focuses on Data Mining and Network Security & Software Engineering based education. she has 2 years of teaching experience and 3 years of Research Experience.

