IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-9, Sept 2019 E-ISSN: 2347-2693

Crop Yield Prediction Based on Data Mining Techniques: A Review

M. Saranya[1*], S. Sathappan[2]

[1]Dept. of Computer Science, Erode Arts & Science College, Erode-9
[2]Dept. of Computer Science, Erode Arts & Science College, Erode-9

*Corresponding Author: drmsaranyamca@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i9.186188 | Available online at: www.ijcseonline.org

Accepted: 12/Sept/2019, Published: 30/Sept/2019

***Abstract*—**Agriculture is the main source of occupation which forms the backbone of our country. It involves the production of crops which may be either food crops or commercial crops. The productivity of crop yield is significantly influenced by various parameters such as rainfall, farm capacity, temperature, crop population density, humidity, irrigation, fertilizer application, solar radiation, type of soil, depth, tillage and soil organic matter. An accurate crop yield prediction support decision-makers in the agriculture sector to predict the yield effectively. Machine learning techniques and deep learning techniques play a significant role in the analysis of data for crop yield prediction. However, the selection of appropriate techniques from the pool of available techniques imposes challenges to the researchers concerning the chosen crop. In this paper, an analysis has been performed on various deep learning and machine learning techniques. To know the limitations of each technique, a comparative analysis is carried out in this paper. In addition to this, a suggestion is provided to further improve the performance of crop yield prediction.

***Keywords*—** Agriculture, crop yield prediction, productivity of crop yield, machine learning, deep learning.

I. INTRODUCTION

Agriculture is the backbone of India as well as the entire world. The development of agricultural technologies is occupied millions of years back and its advancement has been encouraged and described through different environments, civilizations, and technologies. The largest agricultural productivity with the highest quality is achieved by recent prediction techniques that provide the appropriate growth environment under reproduction conveniences protection circumstances. The modern growth of agricultural activities is monitored by data mining techniques and enhanced communication and information processes [1]. Generally, data mining is defined as the process of realizing samples from large databases. The main objective of data mining techniques is to mine the most important information from the database and convert the information into a reasonable formation for additional utilize.

Based on the agricultural applications, the data mining techniques [2] in agriculture are varied. Few data mining techniques are used to forecast environmental conditions such as weather conditions, environmental pollutions and, etc. Few more data mining techniques are used for recognizing purposes such as weed detection, soil characteristics and also used to monitor water cores. Even though these techniques are helpful to predict crop yield, crop yield prediction using data mining techniques is the most significant issue in agriculture. The machine learning technique is also used for crop yield prediction. Machine learning is an application of artificial intelligence that provides system ability to automatically learn and improve from experience. In this paper, an overview of previous researchers using supervised machine learning, unsupervised machine learning, and deep learning is presented. The main aim of this paper is to study detailed information about the previous crop yield prediction using supervised machine learning, unsupervised machine learning, and deep learning techniques. Moreover, their limitations are addressed to further improve the crop yield prediction.

The rest of the paper is structured as follows: Section II presents the analysis of different machine learning and deep learning techniques for crop yield prediction. Section III presents the challenges in machine learning and deep learning technique-based crop yield prediction. Section IV presents the conclusion of this survey.

II. SURVEY ON CROP YIELD PREDICTION TECHNIQUES

A. Crop Yield Prediction using Supervised Machine Learning Techniques

A decision tree algorithm [3] was introduced for soybean yield prediction. The decision tree induction technique was adopted to develop innovative approaches to predict the influence of climatic parameters on the predominant crop productivity of the Bhopal district. A decision tree is a treelike structure where each branch represented an outcome of the test and node represented classes or class distributions. In order to classify an unknown sample, the attribute value of the samples was tested against the decision tree. The findings of the decision tree were framed into different rules for a better understanding of soybean productivity. The decision tree algorithm predicts the crop yield in advance for market dynamics.

A Nonlinear Autoregressive Neural Network (NARNN) [4] was presented for crop yield forecasting. NARNN was a widely used statistical forecasting model for time series. A feed-forward neural network was used to establish NARNN models in which the conventional function was replaced by several neurons that worked together to implicitly approximate the same functionality as a neural network. In addition to this, Nonlinear Autoregressive with External Input Neural Network (NARXNN) was introduced for crop yield forecasting. It was a modified version of NARNN by adding another relevant time series as an extra input to the forecasting model. The NARNN has high accuracy in crop yield forecasting.

A crop yield prediction approach [5] was proposed using a parameter based adaptive neuro-fuzzy inference system (ANFIS) model. In this approach, the crop yield was derived by using attributes. The Neuro-fuzzy, multiple linear regression and fuzzy logic techniques were used to predict the wheat yield. This prediction was obtained based on different input parameters are biomass, rain, radiation, water and extractable soil. The pre-processing technique was employed to remove the outliers, inconsistency, redundant and missing values in the database. Then, Neuro-fuzzy, multiple linear regression and fuzzy logic were applied to predict the wheat yield. The ANFIS based crop yield prediction approach has low RMSE value.

A smart farming system [6] was proposed to predict the crop yield and suggested the optimal climatic factors to maximize the crop yield. This system used support vector regression, multivariate regression and random forest models to predict the crop yield per acre. Initially, yield and weather data were collected from the United States Department of Agriculture. Then, the support vector regression, multivariate regression, and random forest models were applied on the collected dataset to choose the most suitable temperature and moisture content at which the crop yield was optimal. The smart farming system has less root mean absolute error.

B. Crop Yield Prediction using Unsupervised Machine Learning Techniques

A crop yield prediction method [7] was introduced using the Self-Organizing Map (SOM) and multi-layer feed-forward neural network to predict the accurate crops suitable for different kinds of soil. A pre-processing procedure was applied to the large dataset to improve the exactness of the mining process. It converted the raw data adapted for clustering processes. Then, SOM was applied to soil data which categorized soil data based on its geology. It helped to predict the accurate crops suitable for different kinds of soil. Finally, Multi-layer feed-forward neural network was applied to determine the best match of crops for the given type of soil characteristics after the data set was clustered. This crop yield prediction method demonstrated tolerance to a considerable degree over errors present in the training set.

Density-based clustering [8] technique was introduced for crop yield prediction. This technique tried to split the crop data into non-equal clusters based on the Euclidean distance mathematical model. Several parameters were selected for the number of independent factors. Then, the number of clusters was picked to be divided upon the data. The data points were grouped based on the Euclidean distance between the cluster centre and the data points. The clustered data were used to predict crop yield. This technique is suitable for approximate predictions of crop yield.

Weighted-Self Organizing Map (W-SOM) [9] was introduced for crop yield prediction in the Mysore region. The W-SOM was a combination of SOM and Learning Vector Quantization (LVQ). The prediction accuracy of W-SOM was enhanced by minimizing the Within Class Error (WCE) among the clusters. Hence, it provided a clear idea about suitable crop cultivation in the Mysore region. The W-SOM has effective performance.

A crop yield prediction model [10] was proposed using Fuzzy C Means (FCM) clustering and neural network. The FCM allowed one piece of data to belong to two or more clusters that were based on the minimization of the objective function. FCM calculated the output in terms of the degree of membership. Fuzzy clustering dealt with assigning data points in such a manner so that it got the close relationship of similarity between data points as much as possible. The clustered crop data by using FCM was given as input to the neural network which predicted the yield of a crop at a particular region. This model enhanced the basic leadership in accuracy agribusiness.

C. Crop Yield Prediction using Deep Learning Techniques

The impact of the deep neural network [11] on predicting the application rate of fertilizers for coconut trees was presented. DNN is an artificial intelligence model which considered a more efficient way to predict fertilizers under variety cropping patterns. Initially, soil test reports were collected from different places of Thrissur district, Kerala. The collected data consisted of various parameters such as Organic Carbon (%), Soil reaction (pH), Potassium (Kg /ha) Total Soluble Salts (T.S.S) and Phosphorous (Kg /ha). These parameters were processed by DNN to predict the crop yield. The DNN based crop yield prediction has high accuracy.

A deep learning methodology called deep convolutional neural network [12] was introduced for crop yield prediction. Initially, Normalized Difference Vegetation Index (NDVI) and RGB data were collected from Unmanned Aerial Vehicles (UAVs). Nine crop fields located in the vicinity of the city of Pori was selected for this study. These nine fields divided into overlapping data frames of sizes 10m, 20m and 40m. Then a dedicated holdout test data was then constructed from 15% of shuffled data frames and the remaining data frames were used to train the models with k-fold cross-validation. The test data were given as input to deep CNN which predicted the crop yield. The deep CNN has better performance.

A deep neural network [13] was introduced for crop yield prediction. It found the underlying representation of data without handcrafted input of features. DNN had multiple stacked non-linear layers that converted the raw input data (genotype and environmental information) into a higher and more abstract representation at each stacked layer. More complex features were extracted when the DNN network grown deeper. The DNN learned the non-linear and complex relationship between genes, environmental conditions as well as their interactions from historical data. It made a reasonably accurate prediction of crop yield. The DNN based crop yield prediction has superior prediction accuracy.

III. ANALYSIS OF CROP YIELD PREDICTION TECHNIQUES

This section presents a detail about the demerits of different crop yield prediction techniques whose functional information is discussed in the previous section. Through the review of different crop yield predictions using supervised machine learning techniques, unsupervised machine learning techniques, and deep learning techniques, the following challenges are addressed. The decision tree algorithm requires more storage space. The user must be careful when using either NARNN or NARXNN for crop yield forecasting since there is any inconsistency between the results of training and forecasting. The ANFIS based crop yield prediction has strong computational complexity restrictions which is the major drawback of this method. In a smart farming system, more features would be added to improve the crop yield prediction.

The major disadvantage of SOM based crop yield prediction is SOM relied on pre-defined distance in feature space. The density-based clustering is not much more efficient. The prediction rate of W-SOM could be increased by a combination of parallel layer regression along with a deep belief network strategy. The proper selection of a membership function is more difficult in FCM. The DNN based crop yield prediction approach will be extended to monitor and handle the problems likely to be created by the collection of advanced data from a wider area. In deep CNN, CNN will be trained on a large set of features along with times series data to tune the trained model for accuracy. Although DNN captures parameter and yield interactions, its complex model structure makes it hard to produce testable hypotheses that could potentially provide biological insights.

IV. CONCLUSION

In this article, a detailed review of crop yield prediction based on various techniques was presented. According to this analysis, it is known that the deep neural network [13] based crop yield prediction has better performance than other techniques. However, a major limitation of the deep neural network [13] based crop yield prediction technique is its black box property, which is shared by many machine learning methods. Moreover, the complex model structure of DNN makes it hard to produce testable hypotheses that could potentially provide biological insights. The limitation has to overcome this by looking for more advanced models that are not only more accurate but also more explainable.

REFERENCES

[1] B. Milovic, V. Radojevic, *"Application of data mining in agriculture"*, Bulgarian Journal of Agricultural Science, Vol.**21**, Issue.**1**, pp.**26-34,2015**.

[2] D. Ramesh, B. V. Vardhan, *"Data mining techniques and applications to agricultural yield data"*, International Journal of Advanced Research in Computer and Communication Engineering, Vol.**2**, Issue.**9**, pp.**3477-80, 2013**.

[3] S. Veenadhari, B. Mishra, C. D.Singh, *"Soybean productivity modelling using decision tree algorithms"* International Journal of Computer Applications, Vol.**27**, Issue.**7**, pp.**11-15,2011.**

[4] W. W. Guo, H. Xue, *"Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models*", Mathematical Problems in Engineering, **2014**.

[5] Shastry, H. A. Sanjay, M. Hegde, *"A parameter based ANFIS model for crop yield prediction"* IEEE International Advance Computing Conference (IACC), pp.**253-257**, **2015**.

[6] Shah, A. Dubey, V. Hemnani, D. Gala, D. R. Kalbande, *"Smart Farming System: Crop Yield Prediction Using Regression Techniques*". In Proceedings of International Conference on Wireless Communication, Springer, Singapore, pp. **49-56**, **2018**.

[7] F. N. Ogwueleka, *"Crop growth prediction using self-organizing map and multilayer feed-forward neural network*" American-Eurasian Journal of Sustainable Agriculture, Vol.**5**, Issue.**2**, pp.**168-176**, **2018.**

[8] B.V. Vardhan, D. Ramesh, *"Density based clustering technique on crop yield prediction*", International Journal of Electronics and Electrical Engineering, Vol.**2**, Issue.**1,** pp.**56-59**, **2014.**

[9] P. Mohan, K. K. Patil, *"Weather and Crop Prediction Using Modified Self Organizing Map for Mysore Region*", International Journal of Intelligent Engineering & Systems (IJIES), Vol.**11**, Issue.**2**, pp.**192-199**, **2017**.

[10] A.Verma, A. Jatain, S.Bajaj, *"Crop yield prediction of wheat using Fuzzy C Means clustering and neural network*", International Journal of Applied Engineering Research, Vol.**13**, Issue.**11,** pp.**9816-9821**, **2018**.

[11] S. Suchithra, M. L. Pai, *"Impact of Deep Neural Network on predicting application rate of fertilizers (Focus on coconut trees of Kerala northern coastal plain agro ecological unit)*". International Journal of Pure and Applied Mathematics, Vol.**119**, Issue.**10**, pp.**451-466,2017**.

[12] P. Nevavuori, N. Narra, T. Lipping, *"Crop yield prediction with deep convolutional neural networks"* Computers and Electronics in Agriculture, Vol.**163**, **2019.**

[13] S. Khaki, L. Wang, *"Crop yield prediction using deep neural networks*", Frontiers in plant science, Vol.**10,2019**.

