International Journal of
Environmental Sciences & Natural Resources
ISSN: 2572-1119

Juniper Publishers
key to the Researchers

Short Communication
Volume 25 Issue 3 - August 2020
DOI: 10.19080/IJESNR.2020.25.556161

Int J Environ Sci Nat Res

Machine Learning Models for Corn Yield Prediction: A Survey of Literature

Mohsen Shahhosseini and Guiping Hu*

Department of Industrial and Manufacturing Systems Engineering, Iowa State University, USA

Submission: July 27, 2020; **Published:** August 06, 2020

***Corresponding author:** Guiping Hu, Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, Iowa, USA

Abstract

The ability to predict crop yields enables the timely and effective decision making for crop management, and regional agriculture system planning. The field crop corn is the largest crop in the U.S. and hence significant efforts have been devoted to predicting corn yields through various means. The present survey reviews the studies that used machine learning models and their variations to predict corn yield.

Keywords: Agriculture system planning; Crop management; Environmental data; Deep neural networks; Spatial resolution

Introduction

Agriculture and its related industries contribute significantly to the US economy by providing 11% of total U.S employment, and with $1.05 trillion of U.S. gross domestic product (GDP) in 2017 [1]. Crop yield prediction is of great importance as it can deliver insightful information for improving crop management and subsequently U.S. and global economy. In 2019, corn was considered as the largest produced crop in the U.S. [2] and with the increasing demand of corn throughout the country, predicting corn production is essential. The present survey summarizes multiple well-known studies in predicting corn yield using machine learning (ML) models. We first present the most common data preprocessing tasks performed in the literature, and then provide a brief summary of the developed ML models as well as numerical results.

Data Preprocessing Tasks

The most common data preprocessing tasks done by the literature for corn yield prediction include dealing with yearly increasing corn yield trend, feature selection, imputing missing data, and dealing with different spatial resolutions of environmental data sets (soil and weather).

Corn yield trend

Historical corn yields throughout the country demonstrates an increasing trend. This trend is derived from improved genetics (cultivars), improved management, and other technological advances such as farming equipment. Generally, the yearly trend in the corn yields is addressed with two approaches. The first adds the trend back into the developed model as a linear component [2-7]. On the other hand, some studies use recursive neural network variations that are inherently able to capture the time dependency in the response variable [8,9].

Missing data

The missing data treatment strategies have been dependent on the nature of the developed data sets. Some studies impute the missing data with statistical measures [9,10], whereas some other studies made use of expert knowledge to impute the missing data with data aggregation or removing them from the developed data set [4,7,11].

Spatial aggregations

One of the common issues when developing initial data sets arises due to data ingestion from different sources. Each data set has a different spatial resolution. Hence, an important pre-processing task is spatial aggregation to re-arrange the data resolutions of different data sets. The most common solution undertaken in the literature is to use a statistical average/median of the information of the nearest neighbors to coordinate the spatial resolutions of different data sets [3- 8, 12-16].

Machine Learning Models

Various ML models have been designed to predict corn yields throughout the literature, but generally, they can be categorized into five main groups.

Regression-based models

Assuming a linear relationship between the independent and dependent variables, some studies built linear regression models to predict corn yield [6,16]. Other regression-based models in the literature include stepwise linear regression [14], and linear discriminant analysis (LDA) model [17].

Classification and regression tree models

The use of tree models in the literature has been limited due to the superior performance of tree ensemble models. The most common tree-based model has been M5 prime regression model which is an extension of regression tree model with the possibility of linear regression functions at the nodes [18,19].

Tree ensemble models

Tree ensemble models provided better prediction accuracy with the ability to capture nonlinear patterns. Random forest and extreme gradient boosting (XGBoost) have been used more than other tree ensemble models in the literature [3,20].

Neural network models

Like tree ensemble models, neural networks have the ability to deal with nonlinear patterns as well as presenting decent predictions. Many of the recent studies use variations of neural network models from back-propagation neural networks (BPNN) [10] to deep neural networks (DNN) [5,11-13,15], long short-term memory (LSTM) [8] and convolutional neural network (CNN) (Khaki et al., 2020) models.

General ensemble models

Some studies attempted to combine some ML models in an appropriate way to create superior ensemble of models. The base models can be as simple as regression trees or as complex as deep neural networks [4,7].

Summary of Results

The following table summarizes the studies that used ML models to predict US corn yields along with the numerical results of their best developed model (Table 1).

Table 1:

	Data Years	Forecast Level	Test Set	Developed Models	RMSE (Kg/ha)	RRMSE (%)
Jeong et al. [3]	1984-2013	County	50% of the data split randomly	Random forest*, multiple linear regression	1130	16.70%
Kim & Lee [12]	2004-2014	County	Leave-one-year-out CV out-of-bag samples	Deep neural network*, extremely randomized trees, random forest, support vector machines	756	7.30%
Kuwata & Shiba-saki [13]	2008-2013	County	20% of the data split randomly	Deep neural network*, support vector machines	1142	14.00%
Crane-Droesch [5]	1979-2016	County	Out-of-bag samples	Semiparametric neural network*, parametric neural network, non-parametric neural network	998	13.40%
Peng et al. [6]	1982-2016	National	Forward CV Out-of-bag samples	Multiple linear regression	275	2.80%
Khaki & Wang [11]	2008-2016	County	Part of 2016	Deep neural network*, LASSO, Shallow neural network (SNN), regression tree	875	12%
Kim et al. [15]	2006-2015	County	CV Out-of-bag samples	Deep neural network*, multi-ple adaptive regression spines (MARS), support vector ma-chines, random forest, extremely randomized trees, artificial neural network (ANN)	765	7.90%
Shahhosseini et al. [20]	1983-2016	County	2013-2016	Random forest*, XGBoost, multiple linear regression, Ridge regression, LASSO	1400	13.90%
Jiang et al. [8]	2006-2016	County	2016	Long short-term memory (LSTM)*, LASSO, random forest	870	8.20%
Khaki et al. [9]	1980-2018	County	2018	Convolutional neural network + recurrent neural network (CNN-RNN)*, random forest, deep fully connected neural network (DFNN), LASSO	1107	10.30%
Schwalbert et al. [16]	2008-2017	County	CV Out-of-bag samples	Multiple linear regression	1040	11.00%

How to cite this article: Mohsen S, Guiping H. Machine Learning Models for Corn Yield Prediction: A Survey of Literature. Int J Environ Sci Nat Res. 2020; 25(3): 556161. DOI: 10.19080/IJESNR.2020.25.556161

Shahhosseini et al. [7]	2000-2018	County	2016-2018	Optimized weighted ensemble*, multiple linear regression, LASSO, XGBoost, LightGBM, random forest, stacked generalized ensemble, average ensemble, exponentially weighted ensemble	1138	9.50%

Conclusion

We presented a summary of the studies which use machine learning models to predict corn yields. We explained the most common preprocessing tasks that is done to prepare the data for building machine learning models. The developed ML models throughout the literature were categorized into five general groups and a summary of the studies that attempted to predict U.S. corn yields were presented in this study. Reviewing the studies that used simulation crop models and remote sensors to predict corn yields can be considered as a future research direction.

References

1. United States Department of Agriculture, ERS (2019) What is agriculture's share of the overall U.S. economy?
2. Capehart T, Proper S (2019) Corn is America's Largest Crop in 2019.
3. Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, et al. (2016). Random forests for global and regional crop yield predictions. PLoS One 11(6): e0156571.
4. Cai Y, Moore K, Pellegrini A, Elhaddad A, Lessel J, et al. (2017) Crop yield predictions-high resolution statistical model for intra-season forecasts applied to corn in the US. Paper presented at the 2017 Fall Meeting.
5. Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters 13(11):114003.
6. Peng B, Guan K, Pan M, Li Y (2018) Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield. Geophysical Research Letters 45(18): 9662-9671.
7. Shahhosseini M, Hu G, Archontoulis SV (2020) Forecasting corn yield with machine learning ensembles. arXiv preprint arXiv:2001.09055.
8. Jiang H, Hu H, Zhong R, Xu J, Xu J, et al. (2020) A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. Global Change Biology 26(3): 1754-1766.
9. Khaki S, Wang L, Archontoulis SV (2020) A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science 10: 1750.
10. Panda SS, Ames DP, Panigrahi S (2010) Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sensing 2(3): 673-696.
11. Khaki S, Wang L (2019) Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science 10: 621.
12. Kim N, Lee YW (2016) Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography 34(4): 383-390.
13. Kuwata K, Shibasaki R (2016) Estimating corn yield in the united states with modis evi and machine learning methods. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 3(8): 131-136.
14. Satir O, Berberoglu S (2016) Crop yield prediction under soil salinity using satellite derived vegetation indices. Field Crops Research 192: 134-143.
15. Kim N, Ha KJ, Park NW, Cho J, Hong S, et al. (2019) A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006-2015. ISPRS International Journal of Geo-Information 8(5): 240.
16. Schwalbert R, Amado T, Nieto L, Corassa G, Rice C, et al. (2020) Mid-season county-level corn yield forecast for US Corn Belt integrating satellite imagery and weather variables. Crop Science 60(2): 739-750.
17. Mupangwa W, Chipindu L, Nyagumbo I, Mkuhlani S, Sisito G (2020) Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. SN Applied Sciences 2(5): 952.
18. Marinković B, Crnobarac J, Brdar S, Antić B, Jaćimović G, et al. (2009) Data mining approach for predictive modeling of agricultural yield data. Paper presented at the Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment (BioSense09), Novi Sad, Serbia.
19. González Sánchez A, Frausto Solís J, Ojeda Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction. Spanish Journal of Agricultural 12(2).
20. Shahhosseini M, Martinez-Feria RA, Hu G, Archontoulis SV (2019) Maize yield and nitrate loss prediction with machine learning algorithms. Environmental Research Letters 14(12): 124026.

How to cite this article: Mohsen S, Guiping H. Machine Learning Models for Corn Yield Prediction: A Survey of Literature. Int J Environ Sci Nat Res. 2020; 25(3): 556161. DOI: 10.19080/IJESNR.2020.25.556161

DOI:10.19080/IJESNR.2020.25.556161