A Survey on Crop Yield Prediction using Machine Learning

Vishal Nathgosavi and Sachin Patil

March 30, 2021

A Survey on Crop Yield Prediction using Machine Learning

First Author – Vishal J.Nathgosavi
Email-vishalnathgosavi7@gmail.com

Second Author – Dr. Sachin S. Patil
Email-sachin.patil@ritindia.edu

Abstract. In a country in the economy of the nation, the agricultural sector plays an important role. Many advances in the field of agriculture have also been implemented in recent years. Serious research is taking place all over the world with the development of IoT, big data technology and machine learnings about how farmers and even the Govt are learning how to do so. While the fieldwork of agriculture can be classified and analyzed in different areas, such as crop management, ground management, weed identification, water management, yield man- agement, livestock management, we will concentrate on crop yield management because we believe it to be the environment in which a great deal of technical progress is going to allow the farmer to get his crop to his right level. In this article, we will study in particular machine learn- ing and profound learning methods of various researchers as well as their effect in the agricul- ture industry on crop yield management.

Keywords: Precision Agriculture, Crop yield, machine learning, deep learning,

1 Introduction

Although farming work can be classified and analysed in various areas, such as crop management, soil management, weed identification, water management, yield management, farm farming management, we will focus on crop yield management, as we believe that this is the environment that will enable farmers to take advantage of their crops. This paper focuses on computer and in-depth learning methods for different scientists and the effect on crop yield management in the farming industry.[1]. Although farming is classifiable and analysable in different areas such as crop management, soil management, weed identification, water management, crop yield management, we will concentrate on crop farming as we believe this is the climate that will enable farmers to benefit from their crops. This paper is about the machine and in- depth learning for various researchers and how crop yield management in the agricul- ture sector is influenced.
The new technical methodologies in the agriculture sector must be incorporated to feed India's growing population. Moreover, farmers need prompt advice on crop productivity to develop propitious strategies to enhance crop yields. Precision agricul-

ture is a technological method to ensure that soil and crops are provided with what they need to achieve maximum production and health. Precious agriculture uses sen- sors to capture real-time field and weather data and to forecast that farmers will make the right decision. The farm deploys small sensors that capture and transmit the data to the datastore node in question[2]. The data gathered are very large, so big data analysis can be used to process it. Big data provide accurate features such as data collection, data transmission and data analysis. This will also favour farmers and eco- nomic development in the field of agriculture. [3].

The data collected are very high and can be processed by big data analyses. Exact functions such as data storage, data transport and data processing are provided by big data. This will also promote agricultural farmers and economic growth. Years after year ML spreads to more and more different fields of agricultural activity, such as plant management, crop production, soil protection, etc. In the past decade, ML and IoT have demonstrated, along with Big Data Analytics, that this technological assis- tance improves improved management of the agricultural sector as a whole[6]. This paper presents an extensive assessment of the use of ML in agricultural prediction control, usually of crop yield. Several related papers that highlight the main features of common ML models have been presented. This divides the composition of the remainder of the document. In Section 2, we shall examine the agricultural industry's crop yield control as a literature study. Section 3 decribes the Machine Learning part and related work of Machine Learning in Agriculture Crop Yield prediction. Section 4 focuses on our proposed methodologies and in section 5 we had our conclusion.

2 Literature Review

One of the most important predictions for precision agriculture, the yield forecast is of high significance to maximise production in the mapping of returns, yield estimates, balancing crop supply with demand and crop management. Examples of ML applica- tions include the work of [7]; an effective, inexpensive, and non-destructive meth-od that counts coffee fruits automatically on branches. The system measures the fruit of the coffee in three categories: harvestable, not harvestable and disrepair. The system also estimated the coffee fruit weight and maturity rate. This work was designed to provide coffee growers with the knowledge that would optimise their economic gains and schedule their farm work. The developers of [8] in which they created a computer vision method for automation during harvest, are also used for yield prediction. Even if these are not evident, the segments of the system and detect occluded cherry boughs with a complete function. The system's primary goal was to reduce manual harvesting and processing requirements. Authors created an early yield mapping method in an- other study[9] to identify untimely green citrus in an outdoor citrus grove. As with all other relative research, the report aimed to provide farmers with yield-specific knowledge to help them optimise the yield and benefit growth. In another study[10], the authors elaborated on ANNs and multi-temporary remote sensing data to evaluate grassland biomass (kg of dry matter /ha / day). In another study[11], another study focusing on yield prediction and in particular on the prediction of wheat yield was

introduced. The method developed used satellite imaging and was fused with soil data for more precise predictions. [12] the authors proposed a process for detecting tomatoes based on EM photographs that were taken by an aerial unmanned vehicle and remotely sensed red green blue (RGB) (UAV). Also, the authors developed a rice stage forecasting system based on SVM and basic geographical details from China's meteorological stations in the[13] work. Another research has proposed a generalised method for predicting agricultural yield[14]. The approach is based on a long-term agronomic data ENN application (1997–2014). The research focuses on regional predictions of farmers' help to prevent price offers and demand imbalance triggered or hastened by crop quality. in Taiwan in particular.
Table 1 summarizes the above papers for the case of yield prediction sub-category

Table 1. Summary of Crop Yield Prediction literature

Article	Crop	Observed Features	Functionality	Algorithms	Results
[4]	Coffee	Forty-two (42) colour features in digital images illustrating coffee fruits	Automatic count of coffee fruits on a coffee branch	SVM	Harvestable: (1) Ripe/overripe: 82.54–87.83% visibility percentage (2) Semi-ripe: 68.25–85.36% visibility percentage Not harvestable: (1) Unripe: 76.91–81.39% visibility percentage
[5]	Cherry	Coloured digital images depicting leaves, branches, cherry fruits, and the background	Detection of cherry branches with full foliage	BM/GNB	89.6% accuracy
[6]	Green citrus	Image features (form 20 _ 20 pixels digital images of unripe green citrus fruits) such as coarseness, contrast, directionality,line-likeness, regularity, roughness, granularity, irregularity, brightness, smoothness, and fineness	Identification of the number of immature green citrus fruit under natural outdoor conditions	SVM	80.4% accuracy
[7]	Grass	Vegetation indices, spectral bands of red and NIR	Estimation of grassland biomass (kg dry matter/ha/day) for two managed grassland farms in Ireland; Moorepark & Grange	ANN/ANFIS	$R_2 = 0.85$ RMSE = 11.07 Grange: $R_2 = 0.76$ RMSE = 15.35
[8]	Wheat	Normalized values of on-line predicted soil parameters and the satellite NDVI	Wheat yield prediction within-field variation	ANN/SNKs	81.65% accuracy
[9]	Tomato	High spatial resolution RGB images	Detection of tomatoes via RGB images captured by UAV	Clustering/EM	Recall: 0.6066 Precision: 0.9191 F-Measure: 0.7308

3 Machine Learning in Agriculture

To learn from the "experience" (training data), an ML solution typically involves a learning project. A collection of in-stances consists of ML info. Usually, a set of features or variables defines one particular example. The function may be fractional, conditional (e.g., 0 or 1), ordinal (e.g. A+ or B+), or numerical (integer, real, etc.). An optimisation process, which is improved over time with observation, evaluates the consistency of the models for a given task. Several computer and numerical models are employed to measure the ML models and algorithms' accuracy. The training data will be used after the lecture, using the knowledge acquired during the course to iden- tify, forecast or group new examples (test data). Typical ML approach as seen in fig- ure 1.

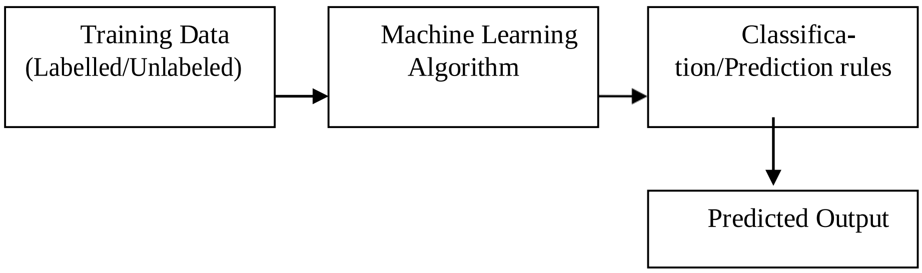

Fig. 1. Machine Learning Process

ML tasks are traditionally divided into many broad categories depending on the type of education (supervised/unsupervised), models of learning (classification, regression, clustering, and dimensional reduction), and learning models used for the chosen task.

3.1 Related work using Machine Learning in Agriculture

A review on the Crop Yield Prediction using Groundnut and Naive Bayes technique in the field of data mining has been completed by Siju, H.L., & Patel, P. J. et al. [6]. Different knowledge mining applications in horticulture is depicted right from the bat and analysed and the study work on the groundnut yield expectation was examined at this stage in conjunction with d Naive Bayes process updated for different applica- tions lastly. Creators also proposed that different knowledge mining techniques can be used to build the more accurate model of groundnut crop yield.

The audits conducted by Chlingaryan, A., Sukkarieh, S., & Whelan, B. et al. [10] focused on AI processes, ROEs and Board precision nitrogen, were discussed. The research reveals the techinque of the value of context proliferation and its consistency in harvest expectations for different vegetation lists. In particular, they show that guassian methods are useful for the prediction and determination of different qualities of plant leaves. The importance of M5-Prime Regression trees is often monitored by creators as a means of identifying various perceptions of yield. Finally, the survey also shows Fuzzy Cognitive Map (FCM), which is to be used for the model and representation of master knowledge for crop yield expectations.

In comparison to a number of information mining techniques like Kmeans and Support Vector Machinery (SVM), and multiple linear regression (MLR) to give high accuracy, K. Samundeeswari, K. Srinivasan et al.[11] was researched in the Krishnagiri District for the dirt information survey. The critical parameter used to increase the development of crops is called the Sun in this article. Creators also included extension and improvement of jobs, using climate conditions and anticipation of harvests for the future.

The various related properties like the location, from which the alkalinity of soil is resolved, are broken down by Bhanumathi, S., Vineeeth, M., & Rohit, N. et al.[12]. In addition, the amount of nitrogen (N), phosphorous (P) and potassium supplements is also considered (K). They used outside applications such as climatic and temperature APIs, soil type, additional dirt assessment and precipitation calculation in that sector to address soil generation.

Nevavuori, P., Narra, N., and Lipping, T et al. [13] have carried out distinctive calculation preparation, including an Adadelta Training algorithm, SGD-force and RMSprop and updated the use of CNNs – an in-depth learning process, through the collection of information using Unmanned Air Vehicles (UAVs), spacecraft and cam- era bundles, and space-filled light. Creators assumed that the RMSprop was mislead- ing and therefore prevented from testing. Adadelta outflanked the SGD power be- tween the two residual calculations and was taken to calculate the readiness for addi- tional exams and This model will be developed in future in order to develop the mod- el for accuracy along with a broader set of highlights like climate and soil.

4.0 Proposed Methodology

Most of the Machine Learning and deep learning algorithms used for crop yield prediction had used very less parameters like only rainfall in that region and area on which the crop production is supposed to be taken. The rainfall and area gets differenti- ated as the district or state gets changed. Although the accuracy of all above models are good and in range of above 95% reaching to 98% but being only 2-3 parameters which we called the features had been identified for analysis and predicting the results for crop yield prediction, but there is still is scope for imprורvemnt in all such models of Artificial intelligence.

Hence we are proposing the machine learning model in which we will work on different features of like rainfall, ph level , temperature, type of soil , precipitaton, Humidity which will give us exact prediction of crop yield in that particular district basaed on all above features. We will implement the various machine learning and deep learning model first to compare and then by analyzing the best model based on performance metrix , we will able to give the best prediction results. Figure 2 respresnts the proposed system architecture.

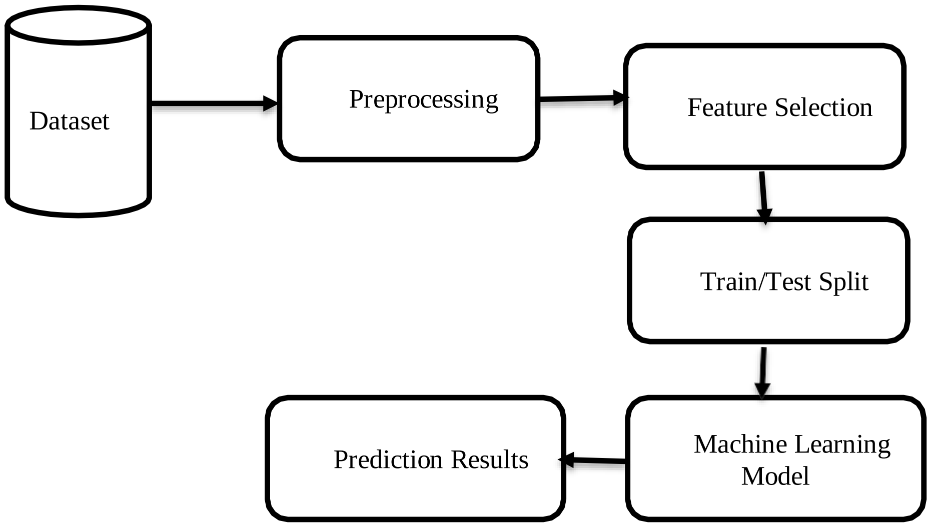

Fig. 2. System Architecture of Proposed System

5.0 Conclusion

We had tried to show that ML models have been applied in multiple applications of agriculture management like crop yield management. This trend in the application distribution reflects the data-intense applications within the crop and high use of images and other data values related to crop yield like rainfall, temp etc. Various ML algorithm like RF, Cubist, soil samples was tested and prediction is done accordingly by various researchers.

It is also evident from the analysis that most of the studies used ANN and SVM ML models. More specifically, ANNs were used mostly for implementations for crop yield management and prediction. This leads the future work for us to take a particu- lar area of the Indian subcontinent and use the available dataset of a region and apply Machine learning and Neural network to predict the crop yield by selecting multiple features so that results will be more appropriate for crop yield prediction.

References

[1] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environ. Dev. Sustain.*, vol. 22, no. 6, pp. 5687–5708, Aug. 2020, doi: 10.1007/s10668-019-00445-x.

[2] A. McBratney, B. Whelan, T. Ancev, and J. Bouma, "Future directions of precision agriculture," in *Precision Agriculture*, Feb. 2005, vol. 6, no. 1, pp. 7–23, doi: 10.1007/s11119-005-0681-8.

[3] D. Howe *et al.*, "Big data: The future of biocuration," *Nature*, vol. 455, no. 7209. Nature Publishing Group, pp. 47–50, Sep. 04, 2008, doi: 10.1038/455047a.

[4] P. J. Ramos, F. A. Prieto, E. C. Montoya, and C. E. Oliveros, "Automatic fruit count on coffee branches using computer vision," *Comput. Electron. Agric.*, vol. 137, pp. 9–22, May 2017, doi: 10.1016/j.compag.2017.03.010.

[5] S. Amatya, M. Karkee, A. Gongal, Q. Zhang, and M. D. Whiting, "Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting," 2015, doi: 10.1016/j.biosystemseng.2015.10.003.

[6] S. Sengupta and W. Lee, "Identification and Determination of the Number of Green Citrus Fruit under Different Ambient Light Conditions," 2012.

[7] I. Ali, F. Cawkwell, E. Dwyer, and S. Green, "Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 7, pp. 3254–3264, Jul. 2017, doi: 10.1109/JSTARS.2016.2561618.

[8] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," 2016, doi: 10.1016/j.compag.2015.11.018.

[9] J. Senthilnath, A. Dokania, M. Kandukuri, R. K.N., G. Anand, and S. N. Omkar, "Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV," *Biosyst. Eng.*, vol. 146, pp. 16–32, 2016, doi: 10.1016/j.biosystemseng.2015.12.003.

[10] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and Electronics in Agriculture*, vol. 151. Elsevier B.V., pp. 61–69, Aug. 01, 2018, doi: 10.1016/j.compag.2018.05.012.

[11] K. Samundeeswari and K. Srinivasan, "CROP YIELD PREDICTION AND SOIL DATA ANALYSIS USING DATA MINING TECHNIQUES IN KRISHNAGIRI DISTRICT," *Int. J. Comput. Sci. Eng.*, vol. 06, no. 08, pp. 49–55, Oct. 2018, doi: 10.26438/ijcse/v6si8.4955.

[12] "Predicting Crop yield and Effective use of Fertilizers using Machine Learning Techniques," doi: 10.35940/ijitee.

[13] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Comput. Electron. Agric.*, vol. 163, Aug. 2019, doi: 10.1016/j.compag.2019.104859.

[14] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74– 78, Oct. 2018, doi: 10.26438/ijcse/v6i10.7478.

[15] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications - A survey," *arXiv*, pp. 1–20, 2019.

[16] L. Shi, Q. Duan, X. Ma, and M. Weng, "The research of support vector machine in agricultural data classification," in *IFIP Advances in Information and Communication Technology*, Oct. 2012, vol. 370 AICT, no. PART 3, pp. 265–269, doi: 10.1007/978-3-642-27275-2_29.