International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 802-806 www.ijaem.net ISSN: 2395-5252

Machine Learning Models for Sustainable Agrarian Applications: A Survey

Zeeshan Choudhary[1], Prof. Sonal Yadav[2], Prof. Pooja Hardia[3]

Date of Submission: 14-06-2023 Date of Acceptance: 24-06-2023

ABSTRACT—Crop yield prediction plays a crucial role in agricultural planning, resource allocation, and food security. In recent years, significant advancements have been made in the field of crop yield prediction, driven by the integration of data-driven approaches, machine learning techniques, and remote sensing technologies. This survey paper aims to provide an overview of the state-of-the-art techniques, challenges, and future directions in the domain of crop yield prediction. We explore various data sources, prediction models, feature extraction methods, and evaluation metrics utilized in crop yield prediction. Additionally, we discuss the challenges faced in the field, including data availability, scalability, and model interpretability. Finally, we identify emerging trends and propose potential directions for future research

Keywords—Crop Yield Prediction, Machine Learning, Agrarian Applications.

I. INTRODUCTION

Agriculture Crop yield prediction is of utmost importance in agriculture and food production for several reasons [1]:

Food Security: Accurate crop yield prediction enables policymakers, farmers, and food organizations to anticipate potential food shortages or surpluses. This information is crucial for ensuring food security by allowing proactive measures such as adjusting import/export policies, implementing effective distribution strategies, and optimizing storage and preservation methods [2].

Resource Allocation: Crop yield prediction helps optimize resource allocation in agriculture, including land, water, fertilizers, and pesticides. By accurately estimating crop yields, farmers can make informed decisions on the allocation of resources, ensuring efficient and sustainable use. This leads to cost reduction, minimized environmental impact, and increased overall productivity[3].

Financial Planning: Reliable crop yield predictions provide valuable insights for farmers and stakeholders in financial planning and risk management. Farmers can anticipate their potential profits, manage loans and investments, and make informed decisions on crop selection and cultivation practices [4]. Accurate predictions help minimize financial risks associated with yield fluctuations and market volatility [5].

Supply Chain Management: Crop yield prediction supports effective supply chain management, facilitating timely harvesting, transportation, processing, and distribution of crops. This ensures that agricultural produce reaches markets and consumers efficiently, minimizing waste and optimizing inventory management [6].

Climate Adaptation: With the increasing impact of climate change on agricultural systems, accurate crop yield prediction aids in climate adaptation strategies. By understanding the potential effects of changing climate patterns on crop productivity, farmers can implement adaptive measures such as altered planting dates, crop selection, and irrigation strategies [7].

Policy and Planning: Crop yield prediction data assists policymakers in formulating evidence-based agricultural policies and strategies. Governments can use this information to design incentives, subsidies, and regulations that promote sustainable farming practices, enhance agricultural productivity, and address regional food security challenges [8].

Research and Development: Crop yield prediction serves as a foundation for agricultural research and development. By analyzing historical yield data and incorporating advanced modeling techniques, scientists can identify key factors influencing crop productivity, study the impact of new technologies, and develop innovative approaches to improve agricultural practices [9]-[10].

Thus, crop yield prediction plays a vital role in ensuring food security, optimizing resource allocation, supporting financial planning, enabling efficient supply chain management, facilitating climate adaptation, guiding policy decisions, and driving agricultural research and development. Accurate predictions contribute to sustainable agriculture, economic stability, and the well-being of both farmers and consumers worldwide [11].

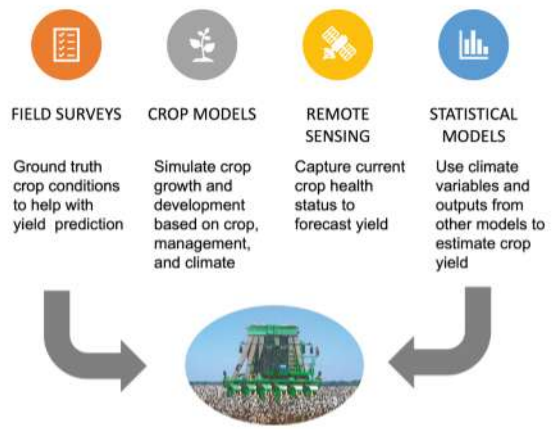

Fig.1 Forecasting Crop Yields
(Source: https://edis.ifas.ufl.edu/publication/AE571)

II. MACHINE LEARNING FOR CROP YIELD PREDICTION

Machine learning models have shown great potential in crop yield prediction by leveraging historical data, environmental factors, and agronomic information. Here are some commonly used machine learning models for crop yield prediction [12]:

Linear Regression: Linear regression is a simple and widely-used machine learning model for crop yield prediction. It establishes a linear relationship between input variables (such as weather conditions, soil properties, and management practices) and crop yield. This model is useful for identifying the impact of individual factors on crop productivity [13].

Decision Trees: Decision tree models, such as Random Forests and Gradient Boosting, are popular for crop yield prediction. Decision trees partition the input space based on specific thresholds and make predictions by averaging multiple decision trees. They can handle nonlinear relationships and capture complex interactions between variables.

Support Vector Machines (SVM): SVM is a powerful model for crop yield prediction that aims to find an optimal hyperplane to separate different classes of data. By using kernel functions, SVM can capture nonlinear relationships in the data. SVM has been successfully applied to crop yield prediction by incorporating various environmental and management factors [14].

Artificial Neural Networks (ANN): ANN models, particularly deep learning architectures like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have shown promising results in crop yield prediction. These models can capture intricate patterns and dependencies within the data, making them suitable for complex agricultural systems [15].

Gaussian Processes (GP): Gaussian Processes are probabilistic models that can provide uncertainty estimates along with yield predictions. GP models are effective in handling limited data and can capture nonlinear relationships between variables. They are useful when uncertainty estimation is crucial, such as in decision-making under risk [16].

Ensemble Methods: Ensemble methods combine multiple models to improve prediction accuracy and stability. Techniques like Bagging, Boosting, and Stacking can be employed with various base models, such as decision trees or neural networks, to enhance crop yield prediction performance [17].

Long Short-Term Memory (LSTM): LSTM is a type of RNN that has shown promise in modeling temporal dependencies and time-series data. LSTM models can capture sequential patterns in historical crop yield data, weather conditions, and other time-dependent factors, enabling accurate prediction of future crop yields [18].

It is important to note that the selection of the appropriate machine learning model depends on the specific characteristics of the dataset, the availability of input variables, and the objectives of the crop yield prediction task. Furthermore, feature engineering, data preprocessing, and model tuning are essential steps in optimizing the performance of machine learning models for crop yield prediction.

III. PREVIOUS WORK

This section highlights the prominent work in the domain.

D. Elavarasan et al. in [1] proposed a a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield. The sequentially stacked layers of Recurrent Neural network is fed by the data parameters. The Q- learning network constructs a crop yield prediction environment based on the input parameters. A linear layer maps the Recurrent Neural Network output values to the Q-values. The reinforcement learning agent incorporates a combination of parametric features with the threshold that assist in predicting crop yield. Finally, the agent receives an aggregate score for the actions performed by minimizing the error and maximizing the forecast accuracy. The proposed model efficiently predicts the crop yield outperforming existing models by preserving the original data distribution with an accuracy of 93.7%.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 802-806 www.ijaem.net ISSN: 2395-5252

C. Dang et al in [2] proposed a Redundancy Analysis (RDA) for feature selection. The simple effects of RDA were used to evaluate the interpretation rates of the explanatory factors. The conditional effects of RDA were adopted to select the features of the explanatory factors. Then, the autumn crop yield was divided into the training set and testing set with an 80/20 ratio, using Support Vector Regression (SVR), Random Forest Regression (RFR), and deep neural network (DNN) for the model, respectively. Finally, the coefficient of determination (R2), the root mean square error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) were used to evaluate the performance of the model comprehensively. The results showed that the interpretation rates of the explanatory factors ranged from 54.3% to 85.0% ($p = 0.002$), which could reflect the autumn crop yields well. When a small number of sample training data (e.g., 80 samples) was used, the DNN model performed better than both SVR and RF models.\

Nigam et al. in [3] proposed a predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc.

Y. Li et al. in [4] proposed a present our statistical modeling practices for predicting rainfed corn yield in the Midwest U.S. and address the aforementioned issues through comprehensive diagnostic analysis. Our results show that vapor pressure deficit and precipitation at a monthly scale, in spline form with customized knots, define the "Best Climate-only" model among alternative climate variables (e.g., air temperature) and fitting functions (e.g., linear or polynomial), with an out-of-sample (leave-one-year-out) median R2 of 0.79 and RMSE of 1.04 t/ha (16.6 bu/acre) from 2003 to 2016.

Bhosale et al. in [5] proposed a data mining approach on agricultural data for yield prediction. It would help in getting results using technologies like data analytics and machine learning and this result will be given to farmers for better crop yield in terms of efficiency and productivity. Authors studied K-means clustering which was used to create clusters. Data stored in clusters facilitated fast search in less time based on cluster hypothesis. The work is expected to help farmers to increase the yield of their crops. Storage of big data in clusters by using K-means clustering algorithm, reduce it to appropriate/valid content using the algorithm. Apriori algorithm helped to count frequently occurring features which helped to predict crop yield for specific location.

T. Islam et al. in [6] proposed an Artificial Neural Network based approach for modeling and prediction of crop yield. This algorithm aims to get better output and prediction, as well as, support vector machine, Logistic Regression, and random forest algorithm is also considered in this study for comparing the accuracy and error rate. Moreover, all of these algorithms used here are just to see how well they performed for a dataset which is over 0.3 million. We have collected 46 parameters such as - maximum and minimum temperature, average rainfall, humidity, climate, weather, and types of land, types of chemical fertilizer, types of soil, soil structure, soil composition, soil moisture, soil consistency, soil reaction and soil texture for applying into this prediction process.

Fernandez et al. in [7] proposed a techniques for the he estimation of yield and total volume of maize production using Spot-5 satellite images and empirical models. These models expressed a) yield (Y) as a function of LAI, and b) yield as a function of NDVI. To determine the efficiency degree of the calculated predictions at the flowering stage of the crop, yield sampling was done at the physiological maturity stage in pilot plots. Regarding yield prediction in the flowering stage, the models Y = f (LAI) reported a value of 5.96 ton.ha-1 and the model Y = f (NDVI) a value of 5.04 ton.ha-1 was obtained. These data represent 114% and 97% respectively of the true yield recorded on the field. The models are specific to the maize crop and the cultivated plots location, and that the forecasts can be acceptably accurate provided the sown areas are precisely determined.

Huang et al. in [8] present a Bayesian model averaging (BMA) method for a multiple crop-growth model ensemble to provide more reliable predictions of maize yields in Liaoning Province, China. The integrated prediction is achieved using a linear combination of the three ensemble members using BMA weights. This integrated approach results in more accurate and precise predictions than any individual model over the entire province. This is because the BMA framework effectively compensates for the uncertainty of individual model simulation and takes advantage of each competing model for reliable prediction.

N. Gandhi et al. in [9] proposed a prediction model based on machine learning. The parameters considered for the present study were precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield for the Kharif season (June to November) for the years 1998 to 2002. The dataset was processed using WEKA tool. A Multilayer Perceptron Neural Network was developed. Cross validation method was used to validate the data. The results showed the accuracy of 97.5% with a sensitivity of 96.3 and specificity of 98.1. Further, mean absolute error, root mean squared error, relative absolute error and root relative squared error were calculated for the present study. The study dataset was also executed using Knowledge Flow of the WEKA tool. The performance of the classifier is visually summarized using ROC curve.

P. Bose et al. in [10] proposed a spiking neural networks (SNNs) for remote sensing spatiotemporal analysis of image time series, which make use of the highly parallel and low-power-consuming neuromorphic hardware platforms possible. This paper illustrates this concept with the introduction of the first SNN computational model for crop yield estimation from normalized difference vegetation index image time series. It presents the development and testing of a methodological framework which utilizes the spatial accumulation of time series of Moderate Resolution Imaging Spectroradiometer 250-m resolution data and historical crop yield data to train an SNN to make timely prediction of crop yield. The research work also includes an analysis on the optimum number of features needed to optimize the results from our experimental data set. The proposed approach was applied to estimate the winter wheat (Triticum aestivum L.) yield in Shandong province, one of the main winter-wheat-growing regions of China.

Research Gap Identified:

1) There may be local fluctuations in the data which need to be filtered out to train a system. The selection of an appropriate mathematical tool in this regard is important.

2) Limited work has been done on use of advanced data pre-processing techniques which can lead to higher predictive performances of classifier with lower computational effort

IV. PERFORMANCE METRICS

Since the purpose of the proposed work is time series prediction, hence it is necessary to compute the required performance metrics. Since there is a chance of positive and negative errors to cancel out, hence it is necessary to compute the Mean Absolute Percentage Error (MAPE) given by:

$$\mathbf{MAPE} = \frac{100}{M}\sum_{t=1}^{N}\frac{E-E_t|}{E_t} \quad (1)$$

Here,
N is the total number of samples
E is the actual value
$\mathbf{E_t}$ is the predicated value

The mean square error is also evaluated often to stop training, which is given mathematically by:

$$\mathbf{MSE} = \frac{1}{N}e_i^2 \quad (2)$$

Here,
E is the error
N is the number of samples

It is always envisaged to attain low error values and high values of accuracy for cloud workload prediction.

CONCLUSION

Accurate crop yield predictions contribute to agricultural research and development. Researchers can analyze historical yield data, climate patterns, and other factors to develop improved crop varieties, pest and disease management strategies, and agronomic practices. By combining historical data, climate information, satellite imagery, machine learning algorithms, and other advanced technologies, crop yield prediction models can provide valuable insights to support farmers, policymakers, and other stakeholders in making informed decisions and optimizing agricultural practices.

REFERENCES

[1]. D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2022, doi: 10.1109/ACCESS.2020.2992480.

[2]. C Dang, Y Liu, H Yue, JX Qian, “Autumn Crop Yield Prediction using Data-Driven Approaches:-Support Vector Machines, Random Forest, and Deep Neural Network Methods”, Canadian Journal of Remote Sensing, Taylor and Franscis, 2021.

[3]. A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125-130, doi: 10.1109/ICIIP47207.2019.8985951.

[4]. Y Li, K Guan, A Yu, B Peng, L Zhao, B Li, J Peng, “Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US”, Field Crops Research Journal, Elsevier 2019, vol. 234, Issue-15 pp. 55-65

[5]. S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid

Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697806.

[6]. T. Islam, T. A. Chisty and A. Chakrabarty, "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh," 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1-6, doi: 10.1109/R10-HTC.2018.8629828.

[7]. Y. M. Fernandez-Ordoñez and J. Soria-Ruiz, "Maize crop yield estimation with remote sensing and empirical models," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 3035-3038, doi: 10.1109/IGARSS.2017.8127638.

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

[9]. N. Gandhi, O. Petkar and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016, pp. 105-110, doi: 10.1109/TIAR.2016.7801222.

[10]. P. Bose, N. K. Kasabov, L. Bruzzone and R. N. Hartono, "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 11, pp. 6563-6573, Nov. 2016, doi: 10.1109/TGRS.2016.2586602.

[11]. T. U. Rehman, S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, ‘‘Current and future applications of statistical machine learning algorithms for agricultural machine vision systems,’’ Comput. Electron. Agricult., vol. 156, pp. 585–605, Jan. 2019.

[12]. D.Elavarasan,D.R.Vincent,V.Sharma,A.Y.Zomaya,andK.Srinivasan, ‘‘Forecasting yield by integrating agrarian factors and machine learning models: A survey,’’ Comput. Electron. Agricult., vol. 155, pp. 257–282, Dec. 2018.

[13]. M. D. Johnson, W. W. Hsieh, A. J. Cannon, A. Davidson, and F. Bédard, ‘‘CropyieldforecastingontheCanadianPrairiesbyremotelysensedvegetation indices and machine learning methods,’’ Agricult. Forest Meteorol., vols. 218–219, pp. 74–84, Mar. 2016.

[14]. A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin, and B. Tekinerdogan, ‘‘Analysis of transfer learning for deep neural network based plant classification models,’’ Comput. Electron. Agricult., vol. 158, pp. 20–29, Mar. 2019.

[15]. A. Kamilaris and F. X. Prenafeta-Boldú, ‘‘Deep learning in agriculture: A survey,’’ Comput. Electron. Agricult., vol. 147, pp. 70–90, Apr. 2018.

[16]. I.M.Evans,‘‘Reinforcement,principle,’’inInternationalEncyclopediaof the Social & Behavioral Sciences, J. D. Wright, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2015, pp. 207–210.

[17]. T Van Klompenburg, A Kassahun, C Catal,‘‘Crop yield prediction using machine learning: A systematic literature review,’’ Computers and Electronics and Agriculture, Elsevier, 2020, vol.177, 105709.

[18]. J. Patel, B. Vala and M. Saiyad, "LSTM-RNN Combined Approach for Crop Yield Prediction On Climatic Constraints," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1477-1483