An Extensive Review of AI Methods for Estimating Crop Yield

Deepesh Agarwal[1], Dharamdas Kumar[2]

Department of mathematics and computer science, Bundelkhand University, Jhansi.

Abstract:
This research aims to investigate the use of Artificial Intelligence (AI) with datasets in precision agriculture, focusing specifically on crop yield estimates. Improvements in remote sensing technologies, especially high-resolution multispectral image datasets, have revolutionized agricultural surveillance. These developments have produced significant insights regarding plant wellness, soil moisture content, and growth patterns of plants. Artificial intelligence systems, which include Machine Learning (ML) as well as Deep Learning (DL) models, can now accurately predict agricultural output by using Vegetation Indices (VIs) obtained from these photos. This study analyzes research from the last five years that uses datasets as well as artificial intelligence methodologies to predict yields for commodities like wheat, maize, and rice. Multiple artificial intelligence strategies are analyzed, including random forest models, support vector machine algorithms , neural networks using convolution (CNNs), and ensemble methods, all of which boost yield forecasts. The analysis underscores a notable disparity in methodological uniformity, as researchers use disparate VIs and AI algorithms for analogous commodities, leading to varied results. This study emphasizes the need of comprehensive comparisons and the use of uniform procedures in future research. The work underscores the vital significance of datasets as well as artificial intelligence in advancing precision agriculture, offering essential insights that future research focused on enhancing both sustainability and effectiveness in crop management using advanced predictive modeling.

Keywords: Crop prediction, smart farming, AI, SVM, CNN.

INTRODUCTION:

The worldwide need for agricultural products and food is rising at a pace commensurate with the continuous expansion of humanity as a whole. This significant increase [1] in demand puts more strain on the Earth's resources. Precision farming (PA) has emerged as an effective strategy to tackle this dilemma and is receiving growing acknowledgment. Since its establishment in 1988, the notion of precise farming has endured for an extended duration. Agricultural production (PA) entails the strategic management of agricultural inputs to enhance crop output while reducing losses.Concerns have emerged about the provision of food which is both safe and nutritious, in light of the rising need for increased agricultural outputs. Monitoring[2] food goods on a worldwide scale is becoming more difficult for national governments as well as consumers. Consequently, tracking and verification have lately become crucial for ensuring the safety of farming operations, leading to a substantial need for geographical data to fulfill these aims.Geographical information is very important for modern farming since it helps with resource management, crop yield improvement, and environmentally friendly methods. Farmers can make smart decisions because they have access to accurate information on a number of factors, such as soil conditions, rainfall, crop health, and the land itself. Producers may utilize geographical data to implement management strategies that are customized to each location. This cuts down on waste and boosts productivity. This data-driven strategy helps farming by encouraging optimal water usage, correct use of herbicides and fertilizers, and early action to prevent pests and diseases.

ESTIMATE OF CROP YIELD AND ARTIFICIAL INTELLIGENCE:

Recently, artificial intelligence has emerged as a primary focus for academics, especially in crop production prediction and forecasting, due to its efficacy and impartiality in processing agricultural data. Yield estimate is an essential component of agriculture, and it is the responsibility of both farmers and agricultural administrators. Accurate yield predictions allow farmers to make educated choices about resource allocation, crop management, and market plan execution. Agriculturists may optimize their use of fertilizers, water, as well as pesticides by forecasting possible crop yields. This not only improves their output but also reduces their environmental effect. Moreover, yield prediction is essential for financial planning, enabling farmers to anticipate their revenue and adeptly handle the risks linked to crop losses or market volatility [3].

By exhibiting this foresight, the farming industry is equipped to efficiently maintain food security economic stability. Consequently, various nations globally are dedicating significant efforts to get yield statistics [4]. RS technologies are increasingly gaining prominence in Pennsylvania because to their very economical nature and less intrusive methodologies [5]. These technologies are designed to predict variability in fields and yields. Artificial intelligence applications may provide precise and instantaneous assessments regarding agricultural yields by using comprehensive information obtained from diverse sources, such as satellite imaging, soil conditions, including crop health metrics. Machine learning algorithms can analyze this information to discern patterns and correlations that traditional approaches may ignore. Additionally, the initiative integrates a novel aspect by providing intelligent agricultural advice. There is a growing interest in the use of artificial intelligence (AI) in farming to provide data-driven insights to growers. Farmers may make educated choices that enhance both production and sustainability by aligning their crop management with precise weather forecasts. The purpose is twofold: to enhance the prediction of weather and to broaden its use in support of modern agriculture methods. This work highlights the significant intersection of environmental science as well as artificial intelligence by providing an approach that is both realistically relevant and widely adaptable across many situations.

BACKGROUND STUDIES:

Given the importance of yield prediction in agriculture and the benefits of artificial intelligence techniques in this field, a considerable body of research focused on AI-based yield forecasting has emerged in recent years. These projects have been thoroughly analyzed in many review studies. Catel C et al. [7] performed an extensive examination of the use of deep learning techniques in predicting agricultural output. They emphasized the significance of using a variety of data sources, including remote sensing photography, among others. The authors emphasized the significant benefits of architectures for deep learning, such as Convolutional Neural Networks, as well as long-short-term memory (LSTM) systems. These algorithms can discern intricate, non-linear correlations among large datasets, resulting in improved accuracy in yield estimates.

Machine learning approaches are often used in agricultural output forecasts. By understanding the intricate linkages among large datasets, these strategies provide a high level of accuracy in forecasting agricultural output. Algorithms often used in agriculture, such as Random Forest (RF) and Support Vector Machine (SVM), have shown successful results, especially when combined with satellite images and vegetation indices like NDVI and EVI. By analyzing data related to agriculture, machine learning improves comprehension of the correlation between crop growth processes and environmental variables, hence facilitating superior decision-making.

Luo et al. [8] investigated the use of machine learning techniques, including support vector machine learning (SVM), Decision Trees (DT), with neural networks, to improve the forecasting of agricultural output and nitrogen levels. They also investigated the capacity of artificial intelligence to improve nitrogen management systems via accurate forecasts of nitrogen levels. Research by Khairunniza-Bejo

and colleagues [9] investigated the use of ANNs, or artificial neural networks, for forecasting agricultural production.The authors highlighted the advantages of artificial neural networks in examining non-linear and multidimensional information in order to evaluate intricate patterns and linkages associated with agricultural development, environmental variables, and management strategies. Dharani et al. [10] focused on the use of deep learning methodologies for predicting crop yields, highlighting the progress made and the efficacy of these methods in agricultural forecasting.Various deep learning models, particularly convolutional neural networks, or CNNs, along with recurrent neural network models (RNNs), have been employed to assess and forecast agricultural outputs employing intricate information such as satellite images and climatic data. They provided an exhaustive summary of previous research concerning these models.
To collect within-field yield of soybeans data, Crusiol et al. [11] used a harvester fitted with a yield monitor throughout 15 separate fields in Brazil, each surpassing 500 hectares in area. They used spectral bands using the dataset and eight unique vegetation indicators to assess yield data using remote sensing. PLSR as well as SVR algorithms were used to construct a connection between the real measurements that the observations derived from various dataset. A 10-fold cross-validation approach was used to evaluate the effectiveness of the PLSR as well as SVR models.

Gómez et al. [12] used satellite imaging and nine machine learning methods to forecast potato production throughout three growth seasons. They persisted in assessing the learning sets that exhibited the greatest performance in the three most recent months. The models were assessed using three years of test data gathered from 35 sites in Spain. The minimum yield sample throughout the three years was 18.47 tons/ha, the highest was 85.79 tons/ha, and the average potato yield was 57.950 tons/ha, with a standard deviation of 16.747. Green potatoes, those with a diameter under 28 mm, malformed potatoes, overall damaged potato were excluded from harvesting and crop production assessments.
Amankulova and coworkers [13] used dataset in a research to predict sunflowers crop yields at both pixel and field scales. Data on crop yields was collected from twenty separate sunflower farms in Hungary in 2020. The scientists used the RF approach to correlate yield readings from an agricultural yield monitoring system coupled with an agricultural combine with NDVI index values collected from dataset wavelengths. Data from 10 fields were used for training, while the other ten fields were allocated for testing.

Deep learning is a method used in AI that is very proficient at managing intricate and multidimensional datasets. In agricultural output forecasting, neural network models like CNN and RNN attain significant accuracy, particularly via the examination of satellite images together with time series data. Deep learning models serve as a potent instrument for improving the understanding of farming operations and forecasting future crop yields, since they can discern intricate and non-linear correlations within extensive datasets. This makes them an exceptionally appropriate tool for agricultural research.

Fernandez-Beltran et al. [14] projected rice production using a dataset. They created RicePAL, an extensive database on Nepali rice agriculture that includes production information from 2016 to 2018. We constructed a three-dimensional convolution neural network for yield forecasting. This research examined the influence of various temporal as well as climatic/soil data combinations on the suggested method. Four revised bands (B02–B04 and B08) and standard twenty-meter bands (B05–B07, B8A, B11, and B12) were included. NDVI pictures obtained from these spectral areas were used. The data were evaluated to assess the impact of climatic and soil conditions on rice crop output. Four scenarios were used to assess the CNN approach with machine learning methodologies (LR, RID, SVR, GPR). All assessments indicated that the CNN exceeded other methodologies. The integration of meteorological along with soil data improved the precision of yield forecasting.

Perich et al. [15] assessed agricultural productivity using artificial intelligence methodologies, using combine harvester data gathered from 2017 to 2021, which reflects five years of minor cereal crop production. Producers in Western Switzerland supplied statistics on agricultural yields over five years. Modeling agricultural yields at the pixel level using high-resolution time series data. Four strategies for estimating agricultural yield using Sentinel-2 data were assessed: 'Partial Integration at Peak' including "Smoothed NDVI' based on spectrum indices, 'Four S2 Scenes' using all spectra bands, and pixels-based RNN yield of crops modeling. The RNN surpassed the other techniques in preprocessing, extracting features, and managing noisy data. The incorporation of data spanning all years improved the models. The "four S2 sequences' technique attained an R2 of 0.88 as well as an RMSE of 10.49% during the evaluation of winter wheat.

Mancini et al. [16] predicted durum wheat production using time series data. The FDA as DL models predicted both temporal as well as spectral yield. By eliminating cloud and shadowing data from the photos, they produced uninterrupted spectral sequences for each pixel. Multifunctional PCA Conventional PLSR was used to forecast wheat output according to NDVI/NDRE curves. Furthermore, the CNN architectures VGG16, VGG19, as well as MobileNetv2 were assessed. Image-based predictions, shown by VGG16, had the best accuracy having a root average square error of 0.047 kg/ha. Deep learning algorithms and spectral indices predict organic durum wheat plant area and crop production.

Amankulova et al. [17] evaluate soybean yields via the use of PlanetScope photography. Throughout the soybean cultivation period, researchers obtained S2 Level-2A and PS Level-3 reflectance from the surface data for standardized differentiated vegetation indicators. The extensive data and decisions tree-based pyDMS enhanced both low- as well as high-resolution photos. A soybean yield estimate model using multidimensional information fusion, deep neural networks, as well as machine learning approaches, while considering within-field variability, was assessed for robustness and adaptability. Joint satellite data fusion surpassed individual forecasts in both precision and error reduction. Employed NDVI maps to corroborate vegetation index crop forecasts. Artificial Neural Networks have outperformed conventional agricultural productivity methods. Farming operators and administrators use satellite data to predict yield.

Desloires, Ienco, and Botrel [18] evaluated end-of-season maize yields in 1,164 farms across two U.S. states from 2017 to 2021. Satellite data were used to acquire ERA-5 temperature and spectrum time series pictures. The research omitted B1, B2, B9, as well as B10. Other uncompressed band pixels were interpolated via 10-meter nearest-neighbor resampling. After eliminating cloud as well as shadow-categorized pixels using an SCL map, pictures with above 50% valid pixels were analyzed. These photos were used to ascertain GNDVI, NDRE, and NDWI. Artificial neural networks (ANNs) trained on Sentinel-2 data measured leaf area index (LAI) other land cover class (LCC), all while random forest (RF), the support vector regression (SVR), multi layer perceptron (MLP), XGBoost, and stacked models forecasted crop yield. The algorithm's hyperparameters were determined using the search grid. Calendar with thermal time resampling indicated diverse yield forecasts.Biophysical variables, spectral bands, indices of vegetation, and their combinations predicted production across four situations. Research indicated that thermal time improves the effectiveness of the stacking algorithm.

Between 2019 and 2021, LIU et al. [19] gathered winter wheat production data from 117 farms in Henan Province using combine harvesters. The Landsat 8 VI time series offered an assessment of pre-harvest production. The BRDF correction model enhanced the precision of image comparisons in photos displaying similar bands from both sensors. Significant connections were identified between several datasets and Landsat 8 bands. Bands B2, B3, B4, B8, B8A, B11, and B12 amalgamated vegetation indices from both sensors across many datasets. NDVI, GNDVI, RVI, EVI2, and WDRVI were

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

calculated from all bands resampled to a 10-meter resolution. Forty days before harvest, the expected yield of winter wheat was estimated.

Darra et al. assessed three tomato yields using Sentinel-2 vegetation indicators [19]. Greek growers, in conjunction with agricultural experts, gathered 2021 production data from 108 farms farming three tomato types over 410.10 hectares. Five vegetation indices (NDVI, WDVI, PVI, RVI, SAVI) were calculated using Sentinel-2 spectral bands to estimate agricultural yield. Visuals were assessed and priced using AutoML. Performance is improved by the automated selection of optimum machine learning methods and the development of ensemble models. Findings demonstrated that AutoML effectively predicted both individual and ensemble yields. Optimal efficacy was achieved with the use of SVR and ARD regression ($R2 = 0.67 \pm 0.02$). Research demonstrated that RVI and SAVI effectively forecasted yield.

Notable machine learning algorithms consist of Random Forest. Algorithms for machine learning, including the use of Support Vector Machine (SVM), logistic regression (LR), as well as Partial Least Squares Regression (PLSR), are used as well. In recent times, CNN, RNN, ACNN, XGBoost, while VGG19 have surfaced as notable enhanced deep learning algorithms. This development signifies the starting point of a new epoch in AI-driven modeling of predictions. The table demonstrates substantial disparities in learning methodologies. Certain studies improve model performance by using many machine learning algorithms, while others depend on a single method. Some research have used sophisticated deep learning architectures, like VGG16 and VGG19, to create more complex prediction models. This demonstrates that academics are determining the most efficacious algorithms and indexes for diverse commodities and scenarios.

DISCUSSION AND CONCLUSION:

Researchers may now precisely outline intricate VI-crop yield interactions using powerful machine learning methods and data from satellites. These models exhibit enhanced precision and dependability by the incorporation of weather and soil data. These methodologies have markedly improved production predictions, hence optimizing agricultural management, resource distribution, and food security initiatives. Artificial intelligence has the potential to predict agricultural output; yet, it is constrained by certain restrictions. Concerns about the generalizability of flawed AI models. Numerous research programs are unsuitable for agricultural use due to geographical or governmental constraints. Models based on historical production data that fail to account for contemporary agricultural methods, climatic fluctuations, or environmental circumstances intensify this problem. Variations in crop varieties, maturity phases, and meteorological circumstances diminish the resilience of these models, preventing them from properly encapsulating the complexity and dynamism of agricultural systems.

AI models in this field use Sentinel-2 images. Vegetation indices serve as proxy indicators for plant health and production, however they represent only a fraction of agricultural yield factors. Models may misjudge productivity by overlooking elements such as soil fertility, precipitation, pest infestations, disease occurrences, and changes in agricultural techniques. Weather circumstances, sensor faults, and illumination fluctuations may impair VI accuracy and bring anomalies into AI system data. Artificial intelligence systems need preprocessing, data cleansing, and validation to provide precise predictions.

Data and artificially intelligent models may enhance agricultural progress. These models aid producers in optimizing distribution of resources, eliminating waste, and mitigating environmental impact via more accurate and timely output predictions. Accurate yield projections may enhance the management of water, fertilizers, pesticides, and agricultural assets, which are vital for sustainable agriculture. Advanced AI methodologies, including deep learning and ensembles approaches, facilitate models in improving predictive accuracy while integrating concerns of climatic variability and other elements influencing agricultural resilience. As agriculture is necessitated to produce more food with less

resources, these innovations will guarantee adequate nutrition and sustainability. AI-driven yield forecasting is crucial for attaining the economic and ecological goals of sustainable agriculture.
The research on AI-driven yield prediction is comprehensive and growing. Many yield prediction studies include data and vegetation indexes, which, when combined with artificial intelligence algorithms, have significant potential to enhance agriculture. The forecasting of agricultural production is getting more accurate and complex, with databases and artificial intelligence applications playing a crucial role in this progress. Progress in this field will improve agriculture management and output. Emerging and evolving fields produce a multitude of new researchers. As the preliminary examination in this domain, our work offers significant support to researchers.

REFERENCES:

[1] Airen, S. and Gupta, P., 2021. AIIoT-Enabled Soil Irrigation System. African Diaspora Journal of Mathematics, 24(1).
[2] Luo, X., Xiong, S., Jia, X., Zeng, Y. and Chen, X., 2025. AIoT-Enabled Data Management for Smart Agriculture: A Comprehensive Review on Emerging Technologies. IEEE Access.
[3] Zhang, H.; Zhang, Y.; Liu, K.; Lan, S.; Gao, T.; Li, M. Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. Comput. Electron. Agric. 2023, 213, 108250
[4] Gupta, P. and Kulkarni, N., 2013. An introduction of soft computing approach over hard computing. International Journal of Latest Trends in Engineering and Technology (IJLTET), 3(1), pp.254-258..
[5] Bhumiphan, N.; Nontapon, J.; Kaewplang, S.; Srihanu, N.; Koedsin, W.; Huete, A. Estimation of Rubber Yield Using Sentinel-2 Satellite Data. Sustainability 2023, 15, 7223.
[6] Darra, N.; Espejo-Garcia, B.; Kasimati, A.; Kriezi, O.; Psomiadis, E.; Fountas, S. Can Satellites Predict Yield? Ensemble Machine Learning and Statistical Analysis of Sentinel-2 Imagery for Processing Tomato Yield Prediction. Sensors 2023, 23, 2586. [
[7] Oikonomidis, A.; Catal, C.; Kassahun, A. Deep learning for crop yield prediction: A systematic literature review. N. Z. J. Crop Hortic. Sci. 2023, 51, 1–26.
[8] Luo, L.; Sun, S.; Xue, J.; Gao, Z.; Zhao, J.; Yin, Y.; Gao, F.; Luan, X. Crop yield estimation based on assimilation of crop models and remote sensing data: A systematic evaluation. Agric. Syst. 2023, 210, 103711
[9] Khairunniza-Bejo, S.; Mustaffha, S.; Ismail, W.I.W. Application of artificial neural network in predicting crop yield: A review. J. Food Sci. Eng. 2014, 4, 1.
[10] Dharani, M.K.; Thamilselvan, R.; Natesan, P.; Kalaivaani, P.C.D.; Santhoshkumar, S. Review on Crop Prediction Using Deep Learning Techniques. J. Phys. Conf. Ser. 2021, 1767, 012026.
[11] rusiol, L.G.T.; Sun, L.; Sibaldelli, R.N.R.; Junior, V.F.; Furlaneti, W.X.; Chen, R.; Sun, Z.; Wuyun, D.; Chen, Z.; Nanni, M.R.; et al. Strategies for monitoring within-field soybean yield using Sentinel-2 Vis-NIR-SWIR spectral bands and machine learning regression methods. Precis. Agric. 2022, 23, 1093–1123.
[12] Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. Remote Sens. 2019, 11, 1745.
[13] Amankulova, K.; Farmonov, N.; Mucsi, L. Time-series analysis of Sentinel-2 satellite images for sunflower yield estimation. Smart Agric. Technol. 2023, 3, 100098.
[14] Fernandez-Beltran, R.; Baidar, T.; Kang, J.; Pla, F. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal. Remote Sens. 2021, 13, 1391.
[15] Perich, G.; Turkoglu, M.O.; Graf, L.V.; Wegner, J.D.; Aasen, H.; Walter, A.; Liebisch, F. Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. Field Crops Res. 2023, 292, 108824.

[16] Mancini, A.; Solfanelli, F.; Coviello, L.; Martini, F.M.; Mandolesi, S.; Zanoli, R. Time Series from Sentinel-2 for Organic Durum Wheat Yield Prediction Using Functional Data Analysis and Deep Learning. Agronomy 2024, 14, 109

[17] Amankulova, K.; Farmonov, N.; Abdelsamei, E.; Szatmári, J.; Khan, W.; Zhran, M.; Rustamov, J.; Akhmedov, S.; Sarimsakov, M.; Mucsi, L. A Novel Fusion Method for Soybean Yield Prediction Using Sentinel-2 and PlanetScope Imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2024, 17, 13694–13707.

[18] Desloires, J.; Ienco, D.; Botrel, A. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. Comput. Electron. Agric. 2023, 209, 107807.

[19] Zhang, H.; Zhang, Y.; Liu, K.; Lan, S.; Gao, T.; Li, M. Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. Comput. Electron. Agric. 2023, 213, 108250.

[18] Bhumiphan, N.; Nontapon, J.; Kaewplang, S.; Srihanu, N.; Koedsin, W.; Huete, A. Estimation of Rubber Yield Using Sentinel-2 Satellite Data. Sustainability 2023, 15, 7223.

[19] Darra, N.; Espejo-Garcia, B.; Kasimati, A.; Kriezi, O.; Psomiadis, E.; Fountas, S. Can Satellites Predict Yield? Ensemble Machine Learning and Statistical Analysis of Sentinel-2 Imagery for Processing Tomato Yield Prediction. Sensors 2023, 23, 2586

[20] Gupta, P., Kumar, S., Singh, Y.B., Singh, P., Sharm, S.K. and Rathore, N.K., 2022. The impact of artificial intelligence on renewable energy systems. NeuroQuantology, 20(16), p.5012

[21] Singh, U., Gupta, P. and Shukla, M., 2022. Activity detection and counting people using Mask-RCNN with bidirectional ConvLSTM. Journal of Intelligent & Fuzzy Systems, 43(5), pp.6505-6520

[22] Singh, U, Gupta, P., Shukla, M., Sharma, V., Varma, S. and Sharma, S.K., 2023. Acknowledgment of patient in sense behaviors using bidirectional ConvLSTM. Concurrency and Computation: Practice and Experience, 35(28), p.e7819.

[23] Rathore, P., Gupta, P., Jain, S. and Shrivastava, Y., 2022. A Study of the Automated Vehicle Number Plate Recognition System. i-manager's Journal on Pattern Recognition, 9(2), p.30.

[24] Manurkar, V., Gupta, P., Sharma, G., Singh, U. and Manurkar, N., 2022. A Face Mask Identification System based on the Internet of Things and Machine Learning for Detecting Covid-19. NeuroQuantology, 20(16), p.3930.

[25] Rajput, A., Gupta, P., Ghodeshwar, P., Varma, S., Sharma, K.K. and Singh, U., 2023, June. Study of Cloud Providers (Azure, Amazon, and Oracle) According To Service Availability and Price. In 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN) (pp. 1177-1188). IEEE.