Original Article

Empirical Survey Analysis For Crop Yield Prediction & Identification Of Factors Affecting Yield Gaps

Preeti Saini[1], Bharti Nagpal[2]

[1]NSUT East Campus (Formerly AIACTR), USICT, Guru Gobind Singh Indraprastha University (GGSIPU), Delhi, India.
[2]Assistant Professor, Department of Computer Science and Engineering, NSUT East Campus (Formerly AIACTR), Delhi, India.

***Corresponding Author:** ritika1516@gmail.com

Doi: 10.47750/pnr.2022.13. S05.207

Abstract

About 70% of India's economy is involved in the agriculture sector to live their lives and contributed to the GDP of the country. The Crop yield information along with the environmental change estimate will be useful for the agriculturalist to decide on price policies prior to harvesting the food source. It establishes a requirement for the prediction model, which precisely determines the harvest conditions, crop varieties, and agricultural yield. In literature, numerous crop prediction methods were devised to estimate crop production in the agricultural field & each technique has its potential in terms of yield forecasting. This review article provides a detailed analysis of the utilized approaches in the literature for the prediction of crop production as well as a discussion on the identification of concerns related to the yield gaps of crops. The discussed approaches were classified based on the application of different strategies, such as Machine learning methods, Deep learning methods, Data mining techniques, vegetative indices, fuzzy logic, and hybrid methods. The study was analyzed based on performance metrics, year of publication, datasets employed, software used for experimentation, and performance attained using various methods and highlights the research gaps of the respective method along with the future direction.

Keywords: Agriculture, Food security, crop yield forecasting, Deep learning, Low Yield

1. INTRODUCTION

Due to the expansion in population all over the world, there is an urge to enhance crop production through various crop management practices. But this, type of pressure can result in the degradation of soil and water quality across the globe with irregular excessive application of fertilizers or irrigation methods over the field. Agriculture productivity is based on various external factors such as weather, soil factors, water availability in the region, nutrients in the soil, fertilizer consumption, etc. The prior information about the aforementioned factors in a region can assist farmers in intake decisions to enhance their crop productivity [15]. To cope with environmental pollution, several countries focused on Organic farming practices instead of chemical-based farming practices. The production of crops as per population requirements is quite a challenging task nowadays due to the risk factors imposed by meteorological parameters. Therefore, appropriate alternative suggestions are regularly in demand in this arena by researchers for the management of Crops, market analysis, and financial planning. There are several methods that have been deployed to represent the linear relation between soil variability and the forecasting parameter of Crop Production [17].

The forecasting of crop yield by utilizing early-season information can add to expanding crop creation and ensuring benefits while reducing input assets and natural contamination. Numerous methods were used for data analysis, for instance, Farmer interviews, Surveys, Metrological data, statistical data from government agencies, satellite imagery, sensor data, and Crop growth patterns [18]. The frameworks build-up for regional forecasting of crop production mostly utilized statistical and climatic information of a particular region of previous years in comparison to yield prediction of the current year. In this kind of framework, climatic information can be selected as the main impetus between yearly yield fluctuations and the Crop Growth Cycle. Many of the authors suggested that Yield forecasts can be improved by constraining vegetation indices (VI's) in crop models instead of only considering meteorological factors exclusively [19]. Crop forecasting can be carried out by manual Surveys, and satellite data, examining distant detecting information, and creating programmed organic products including techniques in plantations. Moreover, conducting manual surveys is considered costly, they require data about the harvests to be gathered in the field, for example, surveying the well-being of plants. Besides, it is hard proportional to abstract the information of various districts and nations for agrarian investigations [39].

With the rapid development of technologies, remote sensing dataset is on-demanding these days as they can identify temporal and spatial features of the field. In recent studies, Remote sensing (RS) innovation is a pre-eminent way to monitor crop development using attributes like Leaf Area Indices (LAI), biomass, leaf, and chlorophyll pigments, which are evaluated based on vegetation Indices (VIs). A study proposed by Zhou et al. [50] determined crop product expectation models covering transmittance band proportions such as NIR/GRN and NIR/RED at initializing phase from the field estimations, and the mentioned frameworks have effectively predicted the enormous region of rice production with the aid of satellite imagery [50]. Despite this, several problems are also associated with satellite imagery because of its low capturing wavelength which will result in poor spatial resolution. To overcome the aforementioned issues sensor techniques such as unmanned aerial vehicle has been utilized nowadays for improving the spectral and temporal key points extraction and they are particularly used for precision farming and plant phenotyping purposes [18]. In the studies related to precision farming, a camera has been installed along with a UAV to capture a large view of fields for distance computation [40]. Nowadays, several studies utilized high spatial resolution multispectral imagery derived from MODIS satellites to apprehend each detail of the Crop Cycle [42].

The following paper is organized as follows: Section 2 presents the Methodology for the review process. Section 3 represents Various Crop Yield Prediction methods. Analysis and discussion of these methods are discussed in Section 4. Section 5 explains research gaps and challenges. Section 6 represents RQ's discussion. Section 7 exhibits the conclusion of the study.

2. MATERIALS AND METHODS

The following Research questions were scrutinized by the Figure 1 methodology to carry out this review.

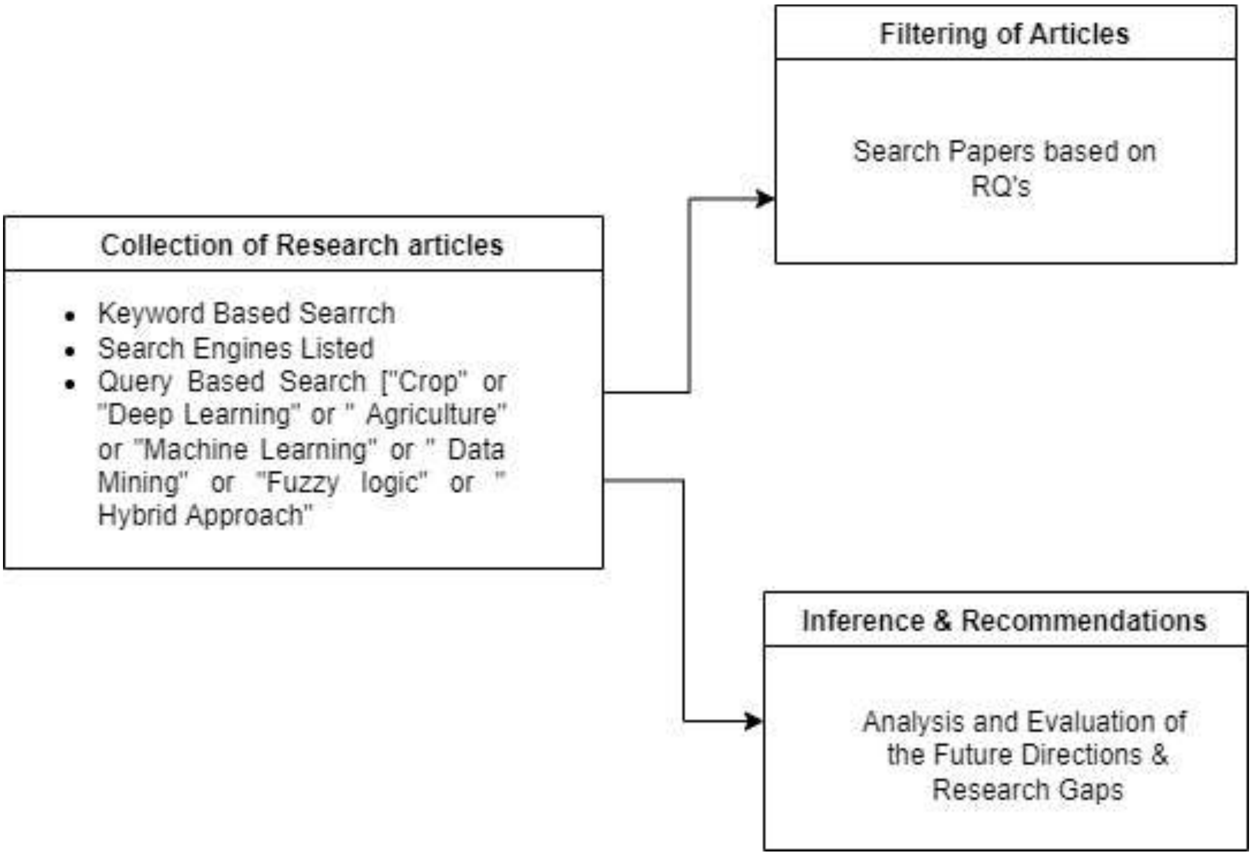

Fig.1. Steps of the Methodology

- ✓ **RQ1.** What were the most commonly vegetative indices utilized in conventional crop product forecasting methods?
- ✓ **RQ2.** Which was the most efficient Classifier in the conventional crop product forecasting model?
- ✓ **RQ3.**Which was the most utilized simulation tool in the conventional crop product forecasting model?
- ✓ **RQ4.** Which were the most often used key metrics for the analysis of the traditional crop yield prediction model?
- ✓ **RQ5.** What were the factors that are mostly responsible for the low yield in regions?

3. REVIEW OF CROP YIELD PREDICTION METHODS

The block diagram for the study of various techniques utilized for crop yield prediction as well as the identification of concerns related to low yield is illustrated in Figure 2.

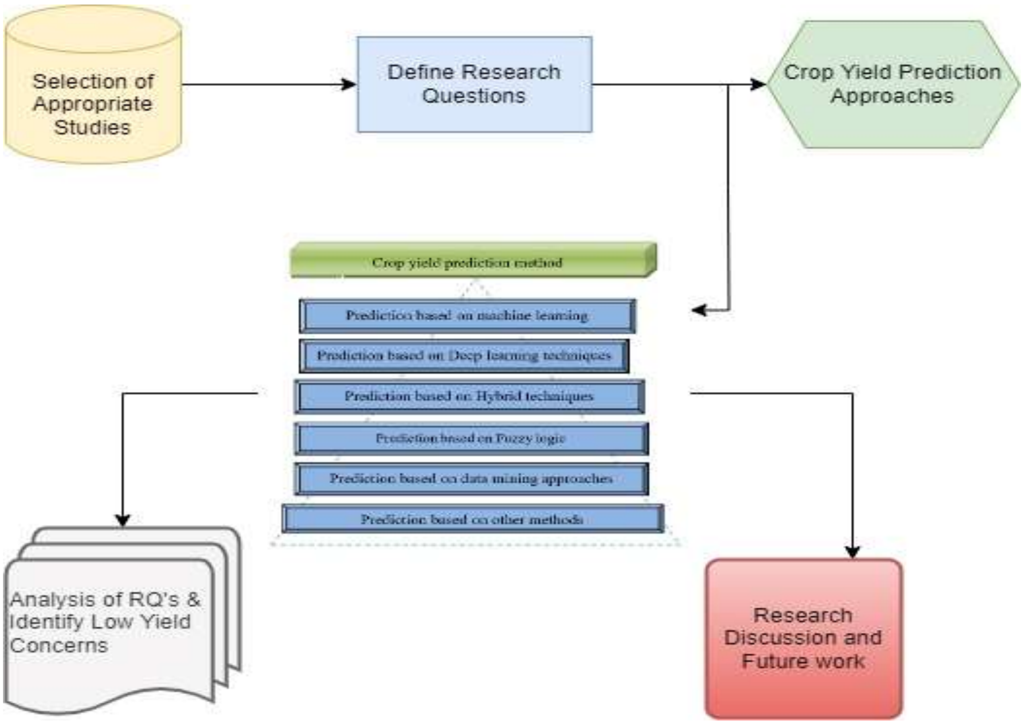

Fig. 2. Block Diagram of the Study

This section is a detailed review of advanced techniques related to the forecasting of crop production illustrates. The advanced methods are classified as machine learning methods, deep learning methods, hybrid techniques, fuzzy logic-based approaches, data mining, vegetative indices, and other methods.
Several techniques were taken into consideration to carry out this study to determine the most efficient feature extractor and classifier for forecasting of production of various Crops despite the several issues. The study of these advanced approaches is further explained in subsections.

3.1. Techniques based on Machine learning

Machine learning is a novel paradigm of Artificial Intelligence that improves results over traditional methods. This section confers about ML techniques referred to in various studies in Table 1.

Table 1. Application of Machine Learning Methods for Crop Yield Prediction

Reference	Technique Used	Database	Results	Tool used
Niketa et al. [4]	Sequential Minimal Optimization method (SMO)	Statistical Data of Maharashtra state	Accuracy= 78%, RMSE=82%	Weka
Sushila et al. [10]	k-means algorithm	Paddy Crop Images	Accuracy=77%	Tensor flow library
Leila Naderloo et al. [29]	adaptive neuro-fuzzy inference system (ANFIS)	Open Interviews	R^2 value = 0.996	MATLAB

3.2. Techniques based on Deep learning

Deep Learning (DL) is a subset of the ML domain and in recent years publishers utilizing this domain of techniques as it provides better results in terms of accuracy despite its black box internal functioning. This section discusses the techniques of DL utilized by various researchers.

Niketa Gandhi et al. [3] demonstrated the enhancement in the rice yield expectation model with the support of ANNs. This methodology has been represented by anticipating rice crop production forecast between the years 1998-2002 for twenty-seven districts of Maharashtra province of India, based on various factors that comprised maximum temperature, precipitation, normal temperature, minimum temperature, harvest evapotranspiration, and yield. The analysis is accomplished with the Weka tool & achieved results in terms of accuracy of 97.5 %. In ANN Multi-Layer Perceptron is utilized to carry out this proposed methodology with an RMSE value is 0.15. Further improvement in results can be done by incorporating a greater number of parameters.

Ekaansh Khosla et al. [8] presented an effective method for predicting the crop yield specifically for the Kharif crops. This model was concentrated on forecasting the summer crops of the coastal locale of Andhra Pradesh, India. In this work rainfall is the most fundamental factor referred to in deciding the measure of summer crop creation, therefore, first, the measure of rainfall is forecasted by utilizing MANN (Modular artificial neural organizations), and afterward, the measure of major summer crops that are produced by utilization of rainfall information of a specific region for particular harvest by Support vector regression method.

Fatin Farhan Haque et al. [13] utilized the Artificial neural organization as the main algorithm for anticipating constant various relapse models. With the current perception, it tends to be reasoned that unsupervised learning is more suitable for obtaining the persistent forecast result than the administered learning algorithm. The current work in ML has generally centered around including investigations where the images are accommodated in the machine vision framework. Information mining agriculture is as yet being chipped away at and this can prompt sorting out the complex method of rural issues.

Abhishek Pandey and Anu Mishra et al. [14] utilized two sorts of ANN such as RBFNN (Radial Basis Function Neural Network) along with GRNN (Generalized Regression Neural Network) to foresee the yield of potato crops, which have been planted unexpectedly at the flat and rough level. The Harvest attributes such as leaf region file, bio-mass, and tallness of herbs were utilized as information for the determination of potato regions as yield information to train and validate the Neural Networks. The GRNN and RBNN anticipated production of potato precisely as it may depend on speedy learning capacity and lower spread consistency (0.5), the GRNN was tracked down as a preferable indicator over RBFNN.

Sivanandhini P et al. [15] predicted the yield of the harvest dependent on various attributes like Humidity, Temperature min, Windspeed, Temperature Max, and Pressure by utilizing neural networks like Recurrent neural networks and Feedforward neural networks. The execution of the neural organization model was assessed via the metric RMSE. On Evaluation proposed work revealed that Feedforward neural network provides the value of RMSE as 0.23 and RNN as 0.00011. Future investigations can be made to upgrade the kind of infections that influence the harvest and recommend utilizing a sort of pesticide to conquer the ailments of the plant.

X.E. Pantazi et al. [17] suggested an adaptive method to foresee field variety in wheat yield, in light of online multi-facet soil information and satellite imagery for crop development qualities. An unsupervised algorithm i.e., Self-Organizing map models such as XY Fusion network, Counter Propagation Artificial Neural Network, and Supervised Kohonen Network was considered and found that among all models SKN provides much more accurate results with 81% as compared to other methods. In this Paper authors used the K-fold Cross-validation method & split the dataset into 75:25 for training & testing datasets.

Maitiniyazi Maimaitijiang et al. [18] evaluated the ability of UAV-based multimodal information in combination with RGB, thermal and multispectral sensors to predict the yield of soybean using the DNN method. All the multispectral, thermal images and the RGB were gathered using a UAV from the testing ground in Columbia of USA. The simulation results revealed that the RGB-based dataset in contrast to spectral data provides more forecast accuracy. Despite of any technique used Multimodal information combination yielded predominant execution for yield expectation over data of a single sensor. It was also found that when the number of features gets expanded then the DNN-based model beat PLSR, SVR, and RFR techniques. Moreover, DNN shows solid flexibility to various assortments with high expectation precision and DNN-based models predict yield forecast model precision: the most noteworthy precision was acquired by DNN-F2 with R^2 0.720 and an overall RMSE value of 15.9%.

Preeti Tiwari and Piyush Shukla [30] utilized the CNN approach with different vegetation indices like SPI, VCI, and NDVI as feature input to forecast the harvest yield. An altered CNN organization with the utilization of a backpropagation algorithm for the error correction method results in high exactness in the anticipated outcomes. An endeavor is made to recommend the best reasonable harvest and its data is dependent on the agricultural area with the current environmental condition that tends to enhance the yield of harvests. The Le Net 5 architecture was utilized as a modified CNN approach and all experiment was carried out using MATLAB. It was stated that the CNN-based forecasting model reduces the relative mistake by 16.17% when contrasted with SNN, while RMSE additionally gets decreased by 16.25%. In the CNN-based forecasting model, the normal processing time for the expectation of harvest yield was diminished by 23.39%.

Yang Chen et al. [32] presented a model known as Crop-SI to forecast the production of the canola, barley, and wheat crop production in the Australian agricultural zone by integrating a radiation-utilized productivity strategy with climatic factors. The study revealed that Harvest SI clarifies 69%, 83%, and 87% of the noticed environmental grain production inconsistency with RMSE of approximately 0.5 for barley, 0.4 for wheat, and 0.4 for canola separately. In this, SI of the crop diminishes the RE in grain production assessment up to 34% for cereal separately, contrasted with two standard frameworks. By fusing water and transposition-driven parameters, the SI of the Crop's prescient expertise in various conditions is improved.

Pritam Bose et al. [33] presented the approach for forecasting the wheat crop before harvesting in the winter season of the Shandong region, China using spiking neural organization of ANN by the employed temporal arrangement of MODIS-NDVI satellite imagery and archival data of studied crop from 2000 to 2013 years. The experimental results showed that the difference between ground truth data and predicted value was 0.13 percent and was able to predict yield forecasting forty days before harvesting. The study concludes with an accuracy rate of 95.64 percent and an error rate of 0.236. The future outlook was provided to extend this work as multiple crops can be sown in one region so there will be a task of recognition of Crop areas. The SNNs perspectives were applied to appraise the winter time of year for wheat production in the Shandong region, particularly in winter-wheat-developing areas of China. This system gave a normal precision of 95.64%, with a predicted error of 0.236 t/ha and a relationship coefficient of 0.801 dependent on a nine-component model. Xiaoqin Dai et al. [38] utilized the two ANN frameworks, one with 6 sources of info (ANN-6) and another ANN framework with 10 sources of info (ANN-10), to reenact the reaction of sunflower concerning the water content with saltiness in soil. In comparison with MLRs, the ANN models provide higher exactness and the RMSEs of 1.6 and 1.1 t ha−1 for ANN-6 and ANN-10, individually, demonstrating that both ANN frameworks can precisely depict the unpredictable connection in the midst of sunflower production, dampness of soil and saltiness at crop phenological stages. Hulya Yalcin et al. [39] utilized the Alex Net architecture of CNN for yield prediction of sunflower Crop fields. The dataset of 500 images was collected by capturing camera images after regular intervals via site TARBIL-mounted camera sets. The result revealed approximately 4% to 7% standard deviation in predicted and ground truth datasets on different fieldsets. The results conclude that the standard deviation value was found to be higher for those images which were not used for training data. The validation set accuracy was found to be 87.67 percent.

Samad Emamgholizadeh et al. [43] utilized neural organizations known as ANN and multiple regression techniques were utilized to forecast the seed production of the sesame Crop. The proposed work considered mentioned parameters such as five plant factors, in particular, blooming season of 100%, plant tallness (PH), container number per plant (CNPP), seed number per case (SNPC) and1000-seed weight (TSW) utilized in the the ANN and MLR to appraise the yield. The two models were investigated on gathered information from the exploration field of Isfahan Educational organization. The outcomes demonstrated that the yield gauges from the ANN were superior to other MLR techniques and demonstrated that the ANN could foresee the SYS more precisely than the traditional MLR and it decreased the MAE and RMSE of SYS assessment by 40.99% and 61.58% contrasted with the MLR technique.

Toshihiro Sakamoto et al. [45] performed a comparative analysis between the traditional approaches such as Linear Regression, Polynomial Regression, and random forest regression methods for yield forecasting of Soya bean & Corn Crops. The time-series connection between the MODIS-wide vegetation file i.e. (WDRVI) of soybean and corn yields was also examined to decide the best and ideal opportunity for recording the MODIS WDRVI as an informative variable in the study region. The Experimentation results showed that the accuracy of the Random Forest method was higher with an RMSE value of 0.539 in comparison to the Polynomial approach having an RMSE of 0.897. It also revealed that the LM strategy, which utilized a linear condition was restricted to past information from the year 2009–2010 & was failed to be adjusted to expand yields experienced as of late without recalibration with the most recent data.

Alexandre Barbosa et al. [47] presented a CNN-based approach to predict Corn yield & experimentation was conducted on Nine on-ranch investigations & compared to conventional methods. The evaluation was performed using R studio software and employed 60:20:20 for training, testing & validation data. The experimental outcomes demonstrate that among all variants of CNN, CNN-LF architecture provides a lower RMSE value, also a decrease in the RMSE up to 68% of the test data when contrasted with different multiple regression and up to 29% when contrasted with the random forest method.

Rashmi Priya et al. [49] presented an effective harvest suggestion framework for the Telangana state by Utilizing the Naive Bayes method using remote sensing & statistical dataset. The framework considered wind speed, temperature, soil, and rainfall attributes for the processing of proposed objectives. The model was evaluated using the Hadoop file system method i.e., Map reduce approach. The predicted results show that the Rice crop can be sown in August-September whereas harvesting can be done during October-November in the studied region**.**

3.3 Techniques Based on Vegetative Index

Several Vegetation Indices are utilized nowadays to extract relevant information from the red & NIR band of remote sensing imagery. Some of the studies are discussed in below-given Table 2.

Table 2. Applications of VI's

Reference	VI's Used	Database	Results	Tool used
Hadi and Farah [11]	EVI & satellite productive estimators	MOD17 Data	R^2 Value as 0.77	N/A
Preeti and Piyush et al. [20]	NDVI, VCI and SPI	Statistical dataset of Madhya Pradesh	RMSE would be decreased to 31.8 percent	MATLAB
Xiaoyang and Qingyuan et al. [31]	Two band of EVI2	AVHRR, MODIS	R^2 value is 0.90	N/A
Yumiko et al. [41]	NDVI	Digitally captured images via Occan optics spectrometer	Study found that impact of water intensity was only impactful if coverage of area is low	R Tool
Vandana et al. [46]	NDVI	MODIS Imagery, questionnaire	R^2 value as 0.69.	N/A
P. Sathiyapriya et al. [48]	NDVI	Climatic dataset	Accuracy=88%	Python

3.4 Techniques Based on Data Mining Strategies

Various studies employed regression, clustering, and classification techniques of Data Mining for forecasting yield prediction. This section highlights some of these studies taken into consideration in Table 3**.**

Table 3. Application of Data Mining Methods for Crop Yield Prediction

Reference	Technique Used	Database	Results	Tool used
Shruti et al. [5]	LAD Tree, IBK, LWL, J48 Classifiers	Maharashtra State Statistical Dataset	80.7% Accuracy	WEKA
Onur and Berberoglu [6]	Stepwise Linear Regression with Radial Basis Function	Landsat Dataset	Mean Percent error (%) for cotton, wheat, and corn were 6.3,7.9,8.8	Python
Yunous et al. [28]	Regression Technique	stochastic periodical temperature profiles	A positive Correlation establish between Temperature & Yield	QGIS
Jordane and Filipe [37]	Regression Models	Statistical Data collected from US Department	Correlation between weather sensitivity & yield =0.79	N/A
Pablo et al. [40]	Multiple Regression	NIR Data of Satellite Imagery	RMSE=79%	N/A
A.B. Payne et al. [44]	Linear Regression	DSLR Camera Farm Images	$R^{2\,=}$ 74%	WEKA

3.5 Techniques Based on Image Processing

In the 21st century, all countries utilize remote sensing data, so there is a huge need for capable image processing techniques to pre-process the acquired data. This section focuses on the image processing studies as illustrated in Table 4.

Table 4. Applications of Image Processing for Yield Forecasting

Reference	Technique Used	Database	Results	Tool used
Dipankar et al. [19]	Polarization Models	Satellite NISAR DATA	R^2 value is 0.73	N/A
Douglas K. and Mark A. [42]	Crop Phenology metric	NASA'S MODIS	R^2 value is 0.73	GIS Tool
X. Zhou et al. [50]	MLR	Remote Sensing data by UAV	R^2 value is 0.79	ENVI

3.6 Techniques Based on Fuzzy Logic

Table 5. Applications of Fuzzy Logic for Yield Forecasting

Reference	Technique Used	Input Parameters	Results
Shivam et al. [9]	ARMAX (Auto Regressive Moving Average with exogenous variable), SARIMA (Seasonal Auto-Regressive Integrated Moving Average), ARMA (Auto Regressive Moving Average)	temperature and rainfall data	RMSE=8.11 & MAE=1.52 for ARMA & 1.03, 1.29 for SARIMA
Akansha et al. [21]	C-means Clustering technique	temperature, solar radiation, and rainfall	Graphical analysis: degree of membership of data values
Bindu et al. [34]	second and third-degree relationship	time-series dataset	a low mean square error was predicted in second-degree fuzzy logic
E.I. Papageorgiou et al. [35]	Fuzzy Cognitive Maps	56 cases (Apple Site)	Prediction rate=75%

3.7 Techniques Based on Hybrid Approach

Shreya V. *et al.* [7] utilized Machine learning techniques with Apriori algorithm and K-means clustering on climate datasets related to a certain region for the prediction of the production output of jowar and linseed Crops. The analysis is done using the tableau tool and different parameters such as rainfall, type of soil, location of Crop region, and Farming area are considered. The Crop Type is determined using the Naïve Bayes method along with the clustering concept to overcome outliers. In this way, the proposed approach assists the farmers in increasing the yield of the crops.

Petteri *et al.* [1] used Convolutional Neural Networks (CNNs), one of the prominent deep learning procedures showing remarkable execution in picture categorization framework that was applied to assemble a framework for crop production expectation dependent on RGB and NDVI information obtained from UAVs. This investigation is a significant advance towards setting up a joined model for grain and wheat yield expectation in the Finnish mainland subarctic environment. The prolonged summer cultivation days in this locale present an extraordinary profile of photoperiod and the temperature, advocating a district explicit profound learning framework for these yields. The outcomes show that the CNN models are fit for sensible exact yield prediction dependent on RGB pictures. Significantly, the CNN engineering appeared to perform preferable with RGB pictures over NDVI pictures. Yet, the attributes such as soil and the environments associated with the time series picture information are needed to be established to tune the prepared model for exactness.

P.S. Maya Gopal and R. Bhargavi *et al.* [2] introduced the characteristic interrelation between ANN and MLR for estimating the yield of the crops. The hybridized MLR-ANN model was presented in this evaluation framework for a proficient harvest yield forecast. The hybridized MLR-ANN model was demonstrated to examine the forecast precision when MLR block and coefficients were utilized to instate the ANN's information layer loads and predisposition. The training algorithm known as backpropagation along with the Feed Forward Artificial Neural Network was utilized for foreseeing precise paddy crop production. In a regular ANN model, the loads and inclination of information and secret layer are initialized arbitrarily. In this amalgamated MLR-ANN framework, rather than irregular loads and inclination introduction, the info layer loads and predisposition are instated by utilizing MLR's coefficients and predisposition. The cross-breed model expectation precision was contrasted with MLR, k-Nearest Neighbor (KNN), ANN, Support Vector Regression (SVR), and Random Forest (RF) models by utilizing execution measurements. The computational time for both crossover MLR-ANN and traditional ANN was calculated. The outcomes show that the proposed mixture MLR-ANN model gives preferred exactness over the customary models.

3.8 Concerns Related to Low-Yield Regions

Various Concerns affected the yield of a crop that includes disease, pests, climate variabilities, biological irregularities, etc. that deteriorate the quality and quantity of a harvest [51]. All over the world, increasing the supply of food increased the fertilizer inputs on farms, will increase soil erosion, and decrease the water level. The factors which are mostly included in the low yield output illustrates in Figure 3.

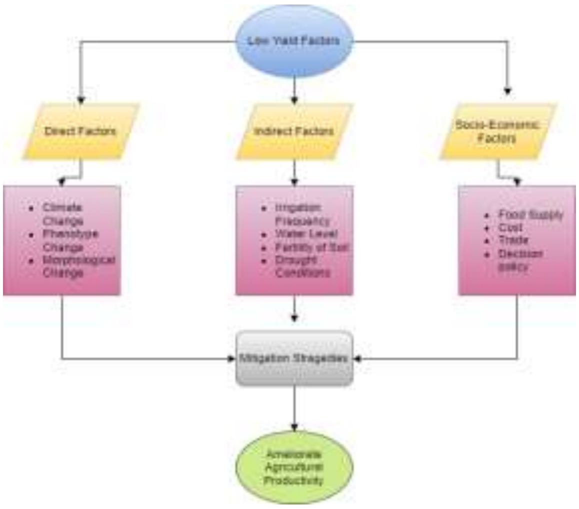

Fig. 3. Factors that Influence Crop Yield

4. EXPERIMENTAL RESULTS

The subsequent section discusses the analysis based on the utilization of the dataset, metrics, techniques of DL used, and frameworks used to implement the proposed architectures in the literature with future directions.

4.1 Analysis based on utilization of various Approaches for Crop Yield Prediction

Figure 4 shows the pictorial representation of different advanced techniques for crop yield prediction utilized in the selected 50 research papers. A Graphical analysis revealed that 30% of publications take advantage of deep learning methods, whereas data mining & machine learning paradigms are used by 17% & 10% of researchers.

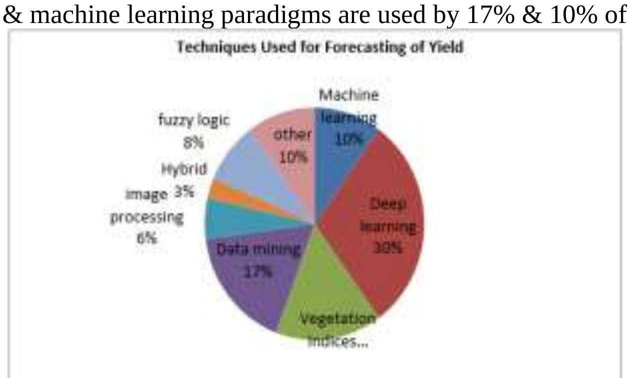

Fig. 4. Category-wise distribution of techniques

4.2 Analysis Based on the Dataset Used

Table 6 illustrates the analysis of the dataset utilized for experimentation in the selected research articles for crop yield prediction. Several studies employed plant village and ImageNet publicly available datasets along with remote sensing derived vegetation indices to capture the spatial information. While others utilized the climatic, soil, Fertilizer related factors for investigation in the studies. Among the 50 papers, 24% used the Normalized vegetation index derived by

MODIS or Landsat satellite imageries, 22% demonstrated the application of weather and soil parameters influence on crop yield prediction approach and others used the publicly or statistically available datasets.
Figure 5 depicts the chart based on the analysis of the dataset.

Table 6. Dataset used in the Studies

Dataset	Research paper
Vegetation Indices data	[1], [14], [17], [25], [30], [33], [41], [42], [46], [47], [48] and [50]
Publicly available data	[2], [3], [4]
Weather and soil data	[5], [7], [8], [15], [16], [23], [24], [43], [44], [45], [49]
Hyper spectral data	[40]
Remote sensing data	[18], [19], [27], [32], [39]
MODIS data	[22], [31], [43]
NASS Statistical data	[37]

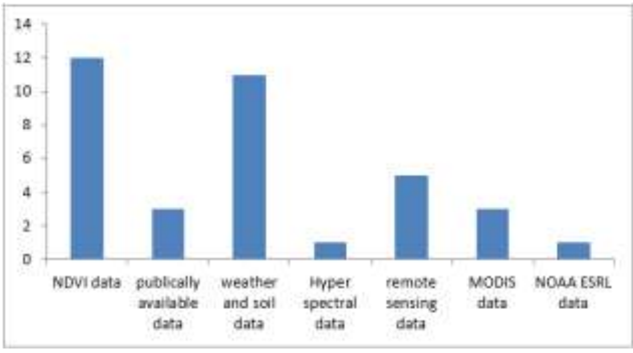

Fig. 5. Analysis based on different datasets

4.3 Analysis Based on Year of Publication

Table 7 exhibits the article related to crop yield domain analysis based on the year of publication. The last 10 years of research work-based papers were considered for analysis, revealing that most of them were published in the years 2018, 2019, and 2020. The papers from the last 10 years were considered for review.

Table 7. Year of publication

Year	Research papers
2010	[24], [25],
2011	[38],
2012	[28], [29]
2013	[22], [26], [35], [42], [44]
2014	[46]
2015	[11], [16], [43]
2016	[3], [4], [6], [12], [17], [31], [33], [40], [41],
2017	[14], [34], [50]
2018	[5], [7], [21], [23], [30], [36], [37], [49]
2019	[1], [2], [18], [20] [8], [9], [10], [39]
2020	[13], [15], [19], [27], [32], [45], [47]
2021	[48]

From the table, it is illustrated that from the 50 papers 2 papers are taken from the year 2010, 1 paper is from 2011, 2 papers are from 2012, 5 papers are from 2013, 1 paper is from 2014, 3 papers are from 2015, 9 papers are from 2016, 3 papers are from 2017, 8 papers are from 2018, 8 papers are from 2019, 7 papers are from 2020 and 1 paper is from 2021. The chart analysis of the year of publication is demonstrated in Figure 6.

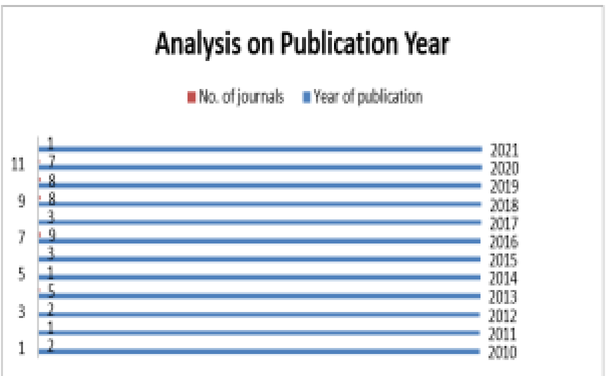

Fig. 6. Analysis based on year of publication.

4.4 Analysis Based on Performance Metrics

Table 8 demonstrates the evaluation of the 50 research papers in terms of performance indices utilized in the crop yield prediction method.

Table 8. Performance metrics used in the Study

Performance metrics	Research papers
MAP	[1], [6], [8], [16]
RMSE	[22], [25], [26], [27], [28], [2], [4], [5], [6], [9], [15], [16], [19], [20], [21], [29]
MBE	[22], [25]
Accuracy	[23], [10], [17]
MSE	[23], [13], [30]
Error rate	[24], [12]
Detection rate	[24]
MAE	[3], [4], [5], [25], [9]
R-square value	[26], [11], [29]
RAE	[4], [5]
RRSE	[4], [16]
MPE	[6]
Average prediction	[7]
MEI	[16]
Prediction Error	[18]
PSNR	[21]
FAR	[21]
FRR	[21]

From the table it is illustrated that the MAPE is considered as the parameter for 4 reviewed research articles, 16 papers utilize the RMSE as their key parameters, and 5 papers utilize MSE as their key parameters. Figure 7 shows the analysis based on the performance metrics.

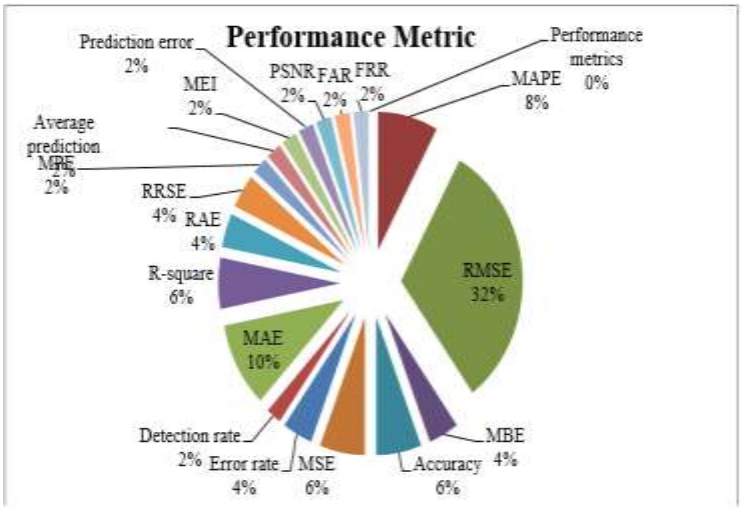

Fig. 7. Analysis based on performance indices

5. RESEARCH GAP AND CHALLENGES

The drawbacks associated with the reviewed strategies from a number of literature are deliberated in this section**.**

5.1 Crop yield prediction using Machine learning methods

a) Shortcomings:

- The most common limitation of machine learning techniques is the inefficient data and complex nature in comparison to traditional techniques.
- It may produce inaccurate results.

b) Future work:

- The future direction includes the selection of appropriate Hyper parameters by incorporating the optimal algorithm that enhances the efficiency of the ML technique.
- To improve accuracy, proper handling of missing values is required in the training dataset of Climate data for Crop yield Forecasting
- Mapping of new features from the existing features of the dataset can be done to enhance the efficiency of models.

5.2 Crop yield prediction using Deep learning methods

a) Shortcomings:

- In Deep learning-based crop yield production methods, the presence of noise in the uncompressed data may be considered the main issue. The noise will affect the performance of the prediction system.
- The huge training time requirement is also an issue experienced in the DL-based crop yield method to process the unstructured data.

b) Future work:

- In the future, Hyperparameters can be optimized with some optimization-based methods that result in the selection of optimal parameters for training the network.
- To improve result outcomes overfitting and underfitting, issues need to be handled.

5.3 Crop yield prediction using Data-mining techniques

a) Shortcomings:

- Gathering on time and a huge amount of data such as soil data and weather data is the main drawback experienced in the crop product forecasting method using data-mining techniques.

b) Future work:

- In the future, standard datasets will be considered to accomplish the intended task of obtaining a better solution.

5.4 Crop yield prediction using VI's approach

a) Shortcomings:

- The major disadvantage of the vegetative index-based method is that quality will be influenced by atmospheric distraction in NIR and red bands of the spectrum.
- The sensitivity of the vegetative indices will intensively increase when the perception of the crop reached a high density.

b) Future work:

- The vegetation indices-based methodology will be improved by combining it with ground truth data for a particular region in analyzing Land usage.
- The sensitivity issue can be reduced by utilizing high-sensor input features.

5.5 Crop yield prediction using a fuzzy logic method

a) Shortcomings:

- The major constraint experienced in the fuzzy logic method is the high degree of complexity, which deteriorates the performance of the crop product forecasting model.
- The lack of additional information about the Coherent nature of the input function of fuzzy integration can reduce the precision of the crop product forecasting model.

b) Future work:

- In the future, highly advanced feature extraction methods will be utilized to extract the highly informative temporal and spatial features to reduce the complexity of the fuzzy logic method.
- The most equipped sensors will be utilized so as to accurately gather the information related to the climatic condition and the soil condition.

6. RESEARCH QUESTION DISCUSSION

- **RQ1-Related (**vegetative indices): The various vegetative indices such as Normalized difference water indices (NDWI), Normalized difference moisture indices (NDMI), and Normalized difference Vegetative indices (NDVI) are utilized in the research. Among the indices, the NDVI was widely utilized vegetative indices in crop yield prediction techniques.

- **RQ2-**Related (classifier): There are many classifiers such as SVM, Naïve Bayes, ANN, and CNN that are utilized for the accurate determination of the crop product forecasting model. The ANN was mostly used in various research as it accurately predicts crop yield.

- **RQ3-**Related (tool); The Weka, which is one of the open-source tools is utilized in the crop product expectation method. The Weka tool was utilized for categorization, training, and clustering in data mining.

- **RQ4-**Related (performance metrics): The Root Square Mean Error is the most widely analyzed key parameter for the evaluation of the conventional method. Almost 16 review articles among 50 papers utilized the RMSE for performance evaluation.

- **RQ5**-Related (Concern to Low Yield): The meteorological, biological, agricultural practices, and irrigation frequency requirements vary globally from one region to another. The studies highlighted that 40% of wheat crop loss occurs due to heat stress in the environment. Similarly, natural calamities such as floods result in excess water to the plants which result in the deterioration of plant physiology. In many countries, these issues are resolved using fertilizer management, rotation system, seed management, and climate management.

7. CONCLUSION

This review article classified the methodologies based on the utilization of strategies and their outcomes, performance metrics, performance attained using various methods, datasets employed, and the year of publication. As per the analysis of the review, it is observed that CNN (Convolutional Neural Network), a technique of deep learning was one of the most utilized DL techniques mentioned in the various research articles. The CNN method followed by the RNN technique of DL provides accurate results and faster execution for effective crop production forecasting. But the main drawback associated with these DL methods is that it requires a huge amount of data for processing, which leads to more requirement of more computational resources. The Weka tool was mostly utilized in almost all the review articles whereas the NDVI was the most accepted vegetative indices in 24% of the literature to predict vegetation cover during the crop season. Approx. 32% of the researchers used the root mean square error as the evaluation metric to predict the accuracy of the proposed model.

This study also highlighted the concerns and solutions related to the low-yield regions that were discussed in the existing articles. As per the Investigation, Climatic factors are the main influence on the outcome of the Yield of various crops in the fields. The appropriate selection of Genomics and crop implementation schemes, and disease management can overcome the issues of low yield of crops. In this way, the future direction includes the selection of appropriate Hyperparameters in the algorithm that enhances the efficiency of Machine learning and Deep learning techniques.

REFERENCES

1. Nevavuori P, Narra N, Lipping T. Crop Yield Prediction with Deep Convolutional Neural Networks. Computers and Electronics in Agriculture. 2019; 163:104859. https://doi.org/10.1016/j.compag.2019.104859 .
2. Gopal PM, Bhargavi R. A Novel Approach for Efficient Crop Yield Prediction. Computers and Electronics in Agriculture.2019;165:104968. https://doi.org/10.1016/j.compag.2019.104968 .
3. Gandhi N, Petkar O, Armstrong LJ. Rice Crop Yield Prediction using Artificial Neural Networks, In proceedings of 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR).2016; 105-110.https://doi.org/10.1109/TIAR.2016.7801222 .
4. Gandhi N, Armstrong LJ, Petkar O, Tripathy AK. Rice Crop Yield Prediction in India using Support Vector Machines, In Proceedings of 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).2016; 1-5. https://doi.org/10.1109/JCSSE.2016.7748856 .
5. Mishra S, Paygude P, Chaudhary S, Idate S. Use of Data Mining in Crop Yield Prediction, In proceedings of IEEE 2nd International Conference on Inventive Systems and Control (ICISC).2018; 796-802. https://doi.org/10.1109/ICISC.2018.8398908 .
6. Satir O, Berberoglu S. Crop Yield Prediction under Soil Salinity using Satellite Derived Vegetation Indices. Field Crops Research.2016; 192:134-143 .https://doi.org/10.1016/j.fcr.2016.04.028 .
7. Bhosale SV, Thombare RA, Dhemey PG, Chaudhari AN. Crop Yield Prediction using Data Analytics and Hybrid Approach. In proceedings of Fourth IEEE International Conference on Computing Communication Control and Automation (ICCUBEA).2018; 1-5.https://doi.org/10.1109/ICCUBEA.2018.8697806 .
8. Khosla E, Dharavath R, Priya R. Crop Yield Prediction using Aggregated Rainfall-based Modular Artificial Neural Networks and Support Vector Regression. Environment, Development and Sustainability.2019; 1-22. https://doi.org/10.1007/s10668-019-00445-x .
9. Bang S, Bishnoi R, Chauhan AS, Dixit AK, Chawla I. Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall Parameters Predicted through ARMA, SARIMA, and ARMAX Models. In proceedings of Twelfth IEEE International Conference on Contemporary Computing (IC3).2019; 1-6. https://doi.org/10.1109/IC3.2019.8844901 .
10. Shidnal S, Latte MV, Kapoor A. Crop Yield Prediction: Two-Tiered Machine Learning Model Approach. International Journal of Information Technology.2019; 1-9.https://doi.org/10.1007/s41870-019-00375-x .
11. Jaafar HH, Ahmad FA. Crop Yield Prediction from Remotely Sensed Vegetation Indices and Primary Productivity in Arid and Semi-Arid Lands. International Journal of Remote Sensing.2015; 36:4570-4589 .https://doi.org/10.1080/01431161.2015.1084434.
12. Gandhi N, Armstrong LJ, Petkar O. Proposed Decision Support System (DSS) for Indian Rice Crop yield Prediction. In proceedings of IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR).2016; 13-18. https://doi.org/10.1109/TIAR.2016.7801205 .
13. Haque FF, Abdelgawad A, Yanambaka VP, Yelamarthi K. Crop Yield Prediction Using Deep Neural Network. In proceedings of IEEE 6th World Forum on Internet of Things (WF-IoT).2020; 1-4.https://doi.org/10.1109/WF-IoT48130.2020.9221298 .
14. Pandey A, Mishra A. Application of Artificial Neural Networks in Yield Prediction of Potato Crop. Russian Agricultural Sciences.2017; 43:266-272.https://doi.org/10.3103/S1068367417030028 .
15. Sivanandhini P, Prakash J. Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network. International Journal of Innovative Science and Research Technology. 2020; 5:1092-1096.
16. Chipanshi A, Zhang Y, Kouadio L, Newlands N, Davidson A, Hill H, Warren R, Qian B, Daneshfar B, Bedard F, Reichert G. Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-Season Prediction of Crop Yield Across the Canadian Agricultural Landscape. Agricultural and Forest Meteorology.2015; 206:137-150.https://doi.org/10.1016/j.agrformet.2015.03.007 .
17. Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM. Wheat Yield Prediction using Machine Learning and Advanced Sensing Techniques. Computers and Electronics in Agriculture.2016; 121:57-65. https://doi.org/10.1016/j.compag.2015.11.018 .
18. Maimaitijiang M, Sagan V, Sidike P, Hartling S, Esposito F, Fritschi FB. Soybean Yield prediction from UAV using Multimodal Data Fusion and Deep learning. Remote Sensing of Environment 2020. 237:111599.https://doi.org/10.1016/j.rse.2019.111599 .
19. Mandal D, Rao YS. SASYA: An Integrated Framework for Crop Biophysical Parameter Retrieval and Within-season Crop Yield Prediction with SAR Remote Sensing data, Remote Sensing Applications: Society and Environment.2020; 20:100366. https://doi.org/10.1016/j.rsase.2020.100366.
20. Tiwari P, Shukla P. Artificial Neural Network-based Crop Yield Prediction using NDVI, SPI, VCI Feature Vectors. In proceedings of Information and Communication Technology for Sustainable Development, Springer, Singapore.2020; 585-594.https://doi.org/10.1007/978-981-13-7166-0_58.
21. Verma A, Jatain A, Bajaj S. Crop Yield Prediction of Wheat using fuzzy c means Clustering and Neural Network. International Journal of Applied Engineering Research.2018; 13:9816-9821.
22. Ines AV, Das NN, Hansen JW, Njoku EG. Assimilation of Remotely Sensed soil Moisture and Vegetation with a Crop Simulation Model for Maize Yield Prediction. Remote Sensing of Environment.2013; 138:149-164.https://doi.org/10.1016/j.rse.2013.07.018 .
23. Islam T, Chisty TA, Chakrabarty A. A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. In proceedings of IEEE Region 10 Humanitarian Technology Conference (R10-HTC).2018; 1-6.https://doi.org/10.1109/R10-HTC.2018.8629828
24. Aggelopoulou AD, Bochtis D, Fountas S, Swain KC, Gemtos TA, Nanos GD. Yield Prediction in apple orchards based on image processing. Precision Agriculture.2011; 12:448-456. https://doi.org/10.1007/s11119-010-9187-0.
25. Mkhabela MS, Bullock P, Raj S, Wang S, Yang Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology.2011; 151:385-393. https://doi.org/10.1016/j.agrformet.2010.11.012

26. Holzman ME, Rivas R, Piccolo MC. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. International Journal of Applied Earth Observation and Geoinformation.2014; 28:181-192. https://doi.org/10.1016/j.jag.2013.12.006 .
27. Sharma S, Rai S, Krishnan NC. Wheat Crop Yield Prediction Using Deep LSTM Model. arXiv preprint 2020. arXiv:2011.01498.
28. Vagh Y, Xiao J. Mining temperature profile data for shire-level crop yield prediction. In proceedings of IEEE International Conference on Machine Learning and Cybernetics.2012; 1: 77-83. https://doi.org/10.1109/ICMLC.2012.6358890 .
29. Naderloo L, Alimardani R, Omid M, Sarmadian F, Javadikia P, Torabi MY, Alimardani F. Application of ANFIS to predict crop yield based on different energy inputs. Measurement.2012; 45: 1406-1413. https://doi.org/10.1016/j.measurement.2012.03.025 .
30. Tiwari P, Shukla P. Crop yield prediction by modified convolutional neural network and geographical indexes. International Journal of Computer Sciences and Engineering.2018; 6:503-513.
31. Zhang X, Zhang Q. Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations. ISPRS Journal of Photogrammetry and Remote Sensing.2016; 114:191-205.https://doi.org/10.1016/j.isprsjprs.2016.02.010 .
32. Chen Y, Donohue RJ, McVicar TR, Waldner F, Mata G, Ota N, Houshmandfar A, Dayal K, Lawes RA. Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. Agricultural and Forest Meteorology.2020; 284:107872. https://doi.org/10.1016/j.agrformet.2019.107872 .
33. Bose P, Kasabov NK, Bruzzone L, Hartono RN (2016) Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. IEEE Transactions on Geoscience and Remote Sensing 54:6563-6573.https://doi.org/10.1109/TGRS.2016.2586602 .
34. Garg B, Aggarwal S, Sokhal J. Crop yield forecasting using fuzzy logic and regression model. Computers & Electrical Engineering.2018; 67:383-403.https://doi.org/10.1016/j.compeleceng.2017.11.015 .
35. Papageorgiou EI, Aggelopoulou KD, Gemtos TA, Nanos GD. Yield prediction in apples using Fuzzy Cognitive Map learning approach. Computers and electronics in agriculture.2013; 91:19-29. https://doi.org/10.1016/j.compag.2012.11.008 .
36. Iizumi T, Shin Y, Kim W, Kim M, Choi J. Global crop yield forecasting using seasonal climate information from a multi-model ensemble. Climate Services.2018; 11:13-23.https://doi.org/10.1016/j.cliser.2018.06.003 .
37. Mathieu JA, Aires F. Assessment of the agro-climatic indices to improve crop yield forecasting. Agricultural and forest meteorology.2018; 253:15-30. https://doi.org/10.1016/j.agrformet.2018.01.031 .
38. Dai X, Huo Z, Wang H. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. Field Crops Research.2018; 121:441-449. https://doi.org/10.1016/j.fcr.2011.01.016 .
39. Yalcin H. An approximation for a relative crop yield estimates from field images using deep learning. In proceedings of 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics).2019; 1-6.https://doi.org/10.1109/Agro-Geoinformatics.2019.8820693 .
40. Rischbeck P, Elsayed S, Mistele B, Barmeier G, Heil K, Schmidhalter U. Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. European Journal of Agronomy.2016; 78:44-59. https://doi.org/10.1016/j.eja.2016.04.013 .
41. Kanke Y, Tubana B, Dalen M, Harrell D. Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. Precision Agriculture.2016; 17:507-530. https://doi.org/10.1007/s11119-016-9433-1 .
42. Bolton DK, Friedl MA. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology.2013; 173:74-84.https://doi.org/10.1016/j.agrformet.2013.01.007 .
43. Emamgholizadeh S, Parsaeian M, Baradaran M. Seed yield prediction of sesame using artificial neural network. European Journal of Agronomy.2013; 68:89-96.
44. Payne AB, Walsh KB, Subedi PP, Jarvis D. Estimation of mango crop yield using image analysis–segmentation method. Computers and electronics in agriculture.2013; 91:57-64.https://doi.org/10.1016/j.eja.2015.04.010 .
45. Sakamoto T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. ISPRS Journal of Photogrammetry and Remote Sensing.2020; 160:208-228. https://doi.org/10.1016/j.isprsjprs.2019.12.012 .
46. Tomar V, Mandal VP, Srivastava P, Patairiya S, Singh K, Ravisankar N, Subash N, Kumar P. Rice equivalent crop yield assessment using MODIS sensors based MOD13A1-NDVI data. IEEE Sensors Journal.2014; 14: 3599-3605.https://doi.org/10.1109/JSEN.2014.2329185 .
47. Barbosa A, Trevisan R, Hovakimyan N, Martin NF. Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture.2020; 170: 105197.https://doi.org/10.1016/j.compag.2019.105197 .
48. Sathiyapriya P, Shobanaa K, Sripriya N, Vaishnavi N. Machine Learning in Crop Yield Prediction Based on Rnn Algorithm. International Research Journal of Engineering and Technology (IRJET) .2021; 08: 118-121.
49. Priya R, Ramesh D, Khosla E. Crop prediction on the region belts of India: A Naïve Bayes MapReduce precision agricultural model. In proceedings of IEEE international conference on advances in computing, communications and informatics (ICACCI). 2018; 99-104. https://doi.org/10.1109/ICACCI.2018.8554948 .
50. Zhou X, Zheng HB, Xu XQ, He JY, Ge XK, Yao X, Cheng T, Zhu Y, Cao WX, Tian YC. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. International Society for Photogrammetry and Remote Sensing Journal.2017; 130:246-255. https://doi.org/10.1016/j.isprsjprs.2017.05.003
51. Liliane TN, and Charles MS. Agronomy - Climate Change & Food Security, Intechopen. 2020. https://doi.org/10.5772/intechopen.90672