Crop Yield Prediction using Machine Learning: A Review of Recent Approaches

Pankaj
Department of CSE,
Sri Venkteshwara University,
J.P. Nagar, India

P.K. Bharti
Department of CSE,
Sri Venkteshwara University,
J.P. Nagar, India

Brajesh Kumar
Department of CSIT,
MJP Rohilkhand University,
Bareilly, India

ABSTRACT
Machine learning is an important tool for the prediction of crop yield. The prediction of yield can help the farmers as well as the policymakers to take timely decisions. With the advanced information on estimated yield, the farmers can make decisions on what to grow to meet the requirement of a growing population. Machine learning techniques can make better yield predictions based on the patterns and correlation information in images or data. There are several machine learning algorithms tested for crop yield prediction. In this work, the recent research works are analyzed in terms of algorithms and the type of information used in prediction studies. It is observed that deep learning techniques have achieved remarkable success in recent times. Most of such methods are based on images of different types such as color, multispectral, and hyperspectral images. This work presents a brief review of the machine learning techniques used for crop yield prediction. The major characteristics and challenges of the methods are discussed and research gaps are identified.

General Terms
Precision agriculture, Machine learling, Phenotyping

Keywords
Crop Yield Prediction; Machine Learning Approaches; Multispectral Images; Algorithmic Classification; Chronological Review

1. INTRODUCTION
India is a country that highly depends on agriculture and a large population depends on agricultural sector for their livelihood. Agriculture can influence the national economic rates greatly. The major objective of the agriculture is to ensure food security for the humans and animals [1]. The agriculture production depends on various factors such as soil, climate, irrigation, fertilizers, seeds, and pests/diseases, etc. The climate change and reduction in agriculture land are badly affecting the agriculture production. Both government and farmers need to make crop yield predictions so that they can make policies accordingly for yield storing, fixing selling rates, importing, selling, and exporting. The policy maker of the country mainly depends on accurate forecasting for performing essential imports as well as exports for enhancing the nation's food protection. Yield forecasting offers more benefits to farmers and cultivators for attaining effective management as well as financial decisions. In agricultural supervision, crop yield analyses are considered indispensable for determining food protection in a region [3]. Forecasting the yield of a crop is considered highly challenging due to several aspects such as genotype, environment, and soil conditions, etc. [2]. Regional factors also affect the crop yield prediction. The traditional yield prediction methods find it difficult to consider all these factors into their prediction model. However, with the advancements in the computation technology, it becomes feasible to develop more accurate prediction models that analyse the larger datasets and relationships between different variables.

The modern machine learning techniques including linear regression, decision tree, random forest, support vector regression, and artificial neural network, etc. treat crop yield as implicit function of weather conditions, soil properties, and crop genotype, etc. However, it is a complex task some linear machine learning methods find it difficult to consider the non-linear relationship between different variables. Modern machine learning techniques such deep learning based models made it possible to use hierarchical structure to learn the complex and non-linear relationship among the variables [3]. The machine learning approaches are more useful for the agronomical crop with accurate predictions [5]. This work aims to review the major machine learning techniques used for crop yield prediction and their evolution in recent years. Also, the strength and weakness of these methods are identified in this work. The major contributions of this work are as follows.

- To provide a review of the existing crop yield prediction methods highlighting major characteristics with their major merits and demerits.
- To provide an algorithmic categorization and performance measures as well as challenges in the conventional yield prediction model.
- To identify research gaps for encouraging the researchers to design improved crop yield prediction models.

The remaining sections of this paper are organized as follows. The categorization of existing machine learning based crop yield prediction methods is given in Section II. A review of literature on recent yield prediction is provided in Section III. The major characteristics, advantages, disadvantages, and challenges in the machine learning models and research gaps discussed in Section IV. Finally, Section V concludes the work.

2. MACHINE LEARNING ALGORITHMS FOR CROP YIELD PREDICTION
A number of machine learning algorithms are used to develop the crop yield prediction models over the years. The parameters such as soil characteristics, weather components, genotype information, and some regional information, etc. are provided as input to these models. The predicted value of the crop yield is obtained as the output of the machine learning model. The major machine learning algorithms used by different researchers for the prediction of crop yield are displayed in Fig. 1. The machine learning methods can be unsupervised or supervised. The supervised algorithms usually perform better than their unsupervised counterparts. This work is therefore concerned with supervised methods only.

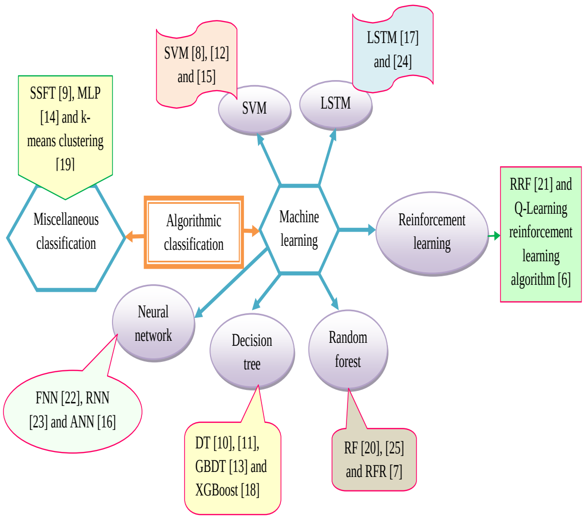

Fig. 1: Classification of Machine Learning Techniques used for Crop Yield Prediction

2.1 Linear regression

Linear regression [29] is a technique that can predict the value of a variable on the basis of the value of another variable by fitting a linear equation to the data. The slop of the best fit line is determined as follows,

$$y = mx + b \tag{1}$$

where m represents the regression coefficient, x is the independent variable, y is the dependent variable, and b is the intercept of the line. It is important to determine the relationship between the two variables of interest before using linear regression. It could be useful only if one variable causes the other. Scatter plots and correlation coefficient are useful tools to find the association between the variables. If there are multiple independent variable to predict the value of a dependent variable then multiple linear regression model is used, which is an extension of the linear regression and Equation (1) becomes

$$y = m_1x_1 + m_2x_2 + \cdots + m_nx_n + b \tag{2}$$

where x terms are input variables and m terms are slop variables. Since crop yield depends on a number of factors, the multiple linear regression is more suitable for the yield prediction.

In its basic form, the linear regression may result in overfitting as it may give more weight to a particular feature. Therefore, some modifications are made in the basic linear regression to overcome this problem. LASSO penalizes the model for sum of absolute weights to reduce the absolute value of weight.

2.2 Support vector regression

Support vector regressor (SVR) [30] is one of the most widely used learning algorithm that can handle non-linear data very well. It is based on the concept of hyperplane, which is a kind of best fit line that helps to predict output values. SVR uses a special kind of function known as kernel that transforms the data in a desired form. It is supervised learning algorithm that uses a training process to determine the hyperplane. However, the performance of SVR highly depends on the value of some hyperparameters. Therefore, it is highly important to choose appropriate values of hyperparameters.

2.3 Decision tree regression

Decision tree [31] uses a tree structured model to make predictions based on the input features. It incrementally develops a decision tree by dividing the dataset into smaller datasets. The topmost decision node in the tree is known as root node. It employs a top-down greedy search algorithm known as ID3 to build the tree from root. The leaf nodes provide the predicted values. Methods like entropy, Gini, information gain, etc. are used measure the prediction error. The weighted average of errors is calculated to provide the final prediction value.

2.5 Random forest regression

The simplicity of decision tree algorithm sometimes underfits the data and decision trees have high variance. The shortcomings of decision tree can be reduced by combining multiple decision trees. The final decision is made on the basis of majority vote. This approach is known as random forest [32] that performs regression by using ensemble learning. The popular ensemble learning techniques are Bagging, Boosting, and Stacking, etc. Random forest uses a Bagging technique that performs calculations in parallel.

2.6 Artificial neural network based regression

Artificial neural network (ANN) [12] is based on human brain that employs a layered architecture to develop learning model. Each layer consists of a number of computing elements known as neurons. The neurons at one layer are connected to the neuron of other neighboring layers. The input data are processed layer by layer and predicted value is provided by the last layer. The number of neurons at input layer depends on the number of features in the data. Usually there are one or more hidden layers that consist of larger number of neurons. Some elementary computations are performed by neurons that collectively solve the complex problems. Appropriate neuron weights are determined during the training process. ANNs are well suited to the non-linear data.

2.7 Deep learning based regression

Deep learning is a sub-branch of machine learning, which is an extension of ANN with more layers. There are various deep learning techniques such deep neural network (DNN), convolutional neural network (CNN), and long short-term memory (LSTM), etc. that have shown promising performance for crop yield prediction on recent years. DNN is similar to ANN having many hidden layers, which are mostly fully connected. CNN [34] on the other hand consists of many different types of layers including convolutional, pooling, and fully-connected. Convolutional layers have weighted neurons that produce feature maps. Pooling layers down sample the feature maps. The fully-connected layer make the predictions. LSTM is an ANN that uses feedback connections that can recall some amount of information from a previous time. LSTMs are especially designed for sequence prediction problems.

3. REVIEW ON CONTRIBUTIONS TO DIVERSE CROP YIELD PREDICTION

Elavarasan and Vincent [6] proposed a combination named recurrent Q-network approach to perform forecasting in the crop yield. The data parameters were offered to the stacked layer presented in the Recurrent Neural Network (RNN). Then, the Q-learning network is developed for crop yield prediction with the help of the input parameter. Finally, agents attained the aggregate score by lowering the error effectively and enhancing the accuracy of the forecast.

Aghighi et al. [7] developed an advanced machine learning approach, that have included different approaches like support vector regression, Boosted Regression Tree (BRT), Gaussian Process Regression (GPR), and Random Forest Regression (RFR), and also their performance were contrasted over existing regression approaches. Then, the outcome attained from several machine learning approaches was averaged for generating better maize yield. Vanli et al. [8] have presented a new approach to the prediction of wheat yield in Turkey. The Top of Atmospheric (TOA) corrections was performed in the entire image. Here, nearly eight machine learning approaches were optimized and tested to perform effective image classification and attained better outcomes in terms of accuracy.

Guo et al. [15] developed a fused framework based on a different combination such as geography data, phenology, and climate for predicting the yield in rice crops with conventional approaches like multiple linear regression and MLR and also with highly advanced machine learning approaches like SVM, RF, and BP. In 2021, Tokhi et al. [16] have presented a yield prediction approach based on machine learning along with time series images. The growing phases of the crop were analyzed by utilizing exponent smoothing approaches in per-pixel value. The seasonality presented in the plants was eliminated by applying an average in the representative mean. ANN approach as well as regression approaches were utilized for the validation and achieved a better accuracy rate. Cao et al. [17] have proposed a new yield prediction model with publically available data in Google Earth Engine (GEE) platform. The developed model attained better outcomes based on the combination of conventional approaches like Long-Short Term Memory (LSTM), RF, and Least Absolute Shrinkage and Selection Operator (LASSO) regression.

Shahhosseini et al. [18] proposed a hybrid machine-learning model offered an effective prediction rate. The developed fused combination provided a highly accurate prediction rate to consider the characteristics of crop modeling with machine learning models. Thus, the developed model attained a better performance rate than another prediction model. Shidnal et al. [19] have presented a neural network-based approach for the classification of several deficiencies that occurred in the plants. Maintaining optimal balance with potassium, phosphorous and nitrogen was highly essential for the crop prediction analysis. The developed model effectively identified the deficiency attained in the plants with the help of images and also predicted the crop yield. In 2019, Filippi et al. [20] have developed a machine-learning model for the prediction of crop yield in an agricultural environment. The attained data based on crop yield were analyzed by utilizing RF approaches. Then, three different approaches were developed according to various conditions such as late-season, pre-sowing, and mid-season for exploring the modification in the prediction capacity.

Elavarasan and Raj [21] developed a combination model based on a regression approach. Here, Reinforcement RF (RRF) was developed to display the improved performance rate with conventional machine learning approaches. The developed technique effectively executed reinforcement learning for the entire selection of separating attributes and achieved an effectively higher outcome in yield prediction. Elavarasan and Raj [22] designed a yield prediction model based on a combination of neural networks. DBN was developed based on the combination of neural network probability and statistics. The DBN approaches effectively performed well in the non-linear systems and also attained better outcomes concerning accuracy, learning speed, and robustness. In 2020, Murali et al. [23] proposed a new combination approach to observe the yield in sugarcane with the help of non-linear data. RNN mostly attained more memory to permit allowable forecasts with a minimal number of parameters. The threshold and weight were optimized with an optimization approach for enhancing the efficacy rate and attained an effectively higher performance rate in terms of accuracy. In 2020, Schwalbert et al. [24] proposed a new approach for performing in-season analysis in forecasting the yield of soybean by utilizing LSTM, weather data, and satellite images. The major goal of the research was to contrast the performance of developed approaches over conventional approaches and attained effectively better performance rate in terms of accuracy.

Seireg et al. [9] designed a combination based on Cascading Regression (CR) and Stacking Regression (SR) to forecast blueberry fields. Here, four different approaches were utilized for the analysis named Sequential Forward Feature Selection (SFFS), extreme gradient boosting based on Feature Importance (XFI), Variance Inflation Factor (VIF), and Sequential Backward Elimination Feature Selection (SBEFS), and attained effective higher yield prediction rate. Pant et al. [10] developed a machine-learning approach for predicting the yield. The machine learning approaches were trained to identify the data patterns effectively for performing crop prediction accurately. Singh et al. [11] proposed a new machine learning framework based on satellite data with a superior resolution rate for predicting the wheat yield. The developed model predicted 36 indicators by utilizing correlation as well as regression approaches.

Son et al. [12] developed a novel approach based on different techniques such as SVM, RF, and ANN. The developed model performed more efficiently than ANN and RF models. The outcome of the developed prediction model achieved an enhanced performance rate in terms of different error metrics. Cheng et al. [13] proposed several indicators to measure the phenological information with the help of machine learning approaches. The developed models effectively validated the yield prediction accuracy rate in the maize field and also effectively detected the optimal time to estimate the yield. Finally, the suggested model has a better robustness rate along with adaptability. Ahmed and Hussain [14] implemented a novel prediction model with machine-learning approaches for estimating the production of wheat. The data related to the crop for five years were collected for analysis and also twelve algorithms were applied to the data set and separated into three sets. The validation outcome was useful for short-listing the best approach for final observation. Ahmed et al. [25] designed a prediction model named Coupled Model Inter-comparison Phase-6 (CMIP6) to analyze the intensity, duration, and magnitude of the crop. Thus, the developed model achieved an enhanced performance rate than the conventional yield prediction model.

Different performance measures achieved and utilized in the conventional yield prediction models are displayed in Table 1. Almost all the existing approaches utilized accuracy and RMSE measure for the prediction of crop yield and also they secured effectively higher performance rates and also other conventional approaches used different performance measures like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Determination Coefficient (DC), Mean Squared Error (MSE), and miscellaneous measures like Median Absolute Error (MedAE), Root Mean Square Percentage Error (RMSPE) and Percentage Absolute Difference (PAD) for the analysis.

4. CHALLENGES IN CROP YIELD PREDICTION

4.1 Characteristics and Challenges

The major features and challenges of the existing crop prediction methods are presented in Table 2. It is observed that amount and quality of data are the major concern in such methods. A number of methods require a good amount of data during phase for providing good prediction accuracy. Apart from that the convergence rate, parameter turning, and overfitting are other issues in machine learning algorithms. Some methods such as one presented in [14] give good accuracy for the data captured under real world conditions but they need a large number for parameters to be tuned to get the desirable results.

Table 1. Analysis on performance measures in crop yield prediction approaches

Citation	Accuracy	RMSE	MAE	MAPE	DC	MSE	Others
[6]	✓	-	-	-	-	-	-
[7]	✓	✓	✓	-	-	-	-
[8]	✓	✓	-	-	-	-	-
[9]	✓	✓	-	-	-	-	-
[10]	✓	-	-	-	-	-	-
[11]	✓	✓	-	-	-	-	-
[12]	-	-	-	✓	-	-	RMSPE
[13]	✓	-	-	-	-	-	-
[14]	-	✓	-	-	-	-	PAD
[15]	✓	-	-	-	-	-	-
[16]	✓	-	-	-	-	-	-
[17]	✓	✓	-	-	-	-	-
[18]	✓	-	-	-	-	-	-
[19]	✓	-	-	-	-	-	-
[20]	✓	-	-	-	-	-	-
[21]	✓	✓	✓	-	✓	✓	-
[22]	✓	✓	✓	-	✓	✓	MedAE
[23]	✓	✓	✓	✓	-	✓	-
[24]	✓	✓	✓	-	✓	-	-
[25]	-	✓	✓	-	-	-	-

Table 2. Advantages and Disadvantages of existing crop yield prediction models

Author name	Citation	Advantage	Disadvantage
Elavarasan and Vin	[6]	It efficiently lowers the dependence on professionals.	It didn't have a better robustness rate.
Aghighi *et al.*	[7]	It has the efficiency to deal with high-dimensional data presented in complex distribution.	It gets suffered when data are used from different geographical areas.
Vanli *et al.*	[8]	It efficiently resolves regression as well as classification issues.	The accuracy of performance highly depends on the data quality.
Seireg *et al.*	[9]	It effectively enhances the accuracy rate and also minimizes the complexity.	It suffered a lot when large datasets are utilized.
Pant *et al*.	[10]	It has better interpretability as well as versatility rate	It faces overfitting issues.
Singh *et al.*	[11]	It effectively enhances the performance of spatial and spectral resolution.	It didn't have a highly accurate accuracy rate.
Son *et al.*	[12]	It has the capability for learning highly complex and non-linear relationships.	It needs more data in the training phase.
Cheng *et al.*	[13]	It has a high robustness rate.	It needs to enhance the spatial adaptability rate effectively.
Ahmed and Hussain	[14]	It can train the data in real-world conditions.	It has an enormous amount of parameters to perform the analysis.
Guo *et al.*	[15]	It accurately predicts the yield in different climate conditions.	It didn't process the data effectively when more noise is presented.
Tokhi *et al.*	[16]	It can eliminate the unknown factor.	It is highly complex to perform analysis and validation.
Cao *et al.*	[17]	It has an effectively high output and input bias rate.	It required more time to train the data and also need more memory.
Shahhosseini *et al*	[18]	It effectively minimizes overfitting issues.	It required more data to enhance the prediction rate.
Shidnal *et al.*	[19]	It offered a better convergence rate.	It suffered a lot due to the variation presented in cluster size.
Filippi *et al.*	[20]	It effectively resolves overfitting issues and enhances the performance rate in terms of accuracy.	It has more imbalanced data and creates more complexity in the system.
Elavarasan and Raj	[21]	It has a fast execution rate and also effectively resolves the overfitting issue.	It didn't concentrate to resolve the error bound.
Elavarasan and Raj	[22]	It performed effectively in a non-linear system and also resolves the gradient diffusion issue.	It didn't have a more robustness rate.
Murali *et al.*	[23]	It can train more input data offered in a different size.	It attained more vulnerability when climate change is attained rapidly.
Schwalbert *et al.*	[24]	It has effectively higher the learning rate and resolves the vanishing issues.	It easily falls into overfitting issues.
Ahmed *et al.*	[25]	It is highly accurate than conventional approaches.	It faces overfitting issues.

4.2 Research Gaps

Agriculture is determined as the major resource area because it offered a huge quantity of food for individuals. In recent days, most nations presented worldwide are highly suffered due to food shortages. Several complexities are attained due to the variability attained in the climatic changes, weather, soil loss, and population enhancement and it is subjected to design an approach to assure the production and growth in a highly reliable manner. Expanding the production of agricultural food is considered highly sustainable for the production of food worldwide. Researchers considered pre-harvest yield prediction approaches as highly important to plan and generate several decisions regarding crop yield. Conventional approaches are termed as costly, consume more time, and are subjective. Designing an accurate prediction model by utilizing weather data faces more complexity. Information technology is utilized commonly for predicting the risk related to agriculture and they are utilized for predicting the yield in crops accurately before harvesting. To perform effective crop yield prediction, different kinds of data like satellite imagery, meteorological data, oil data, agricultural statics, remotely sensed data, high and low spectral as well as spatial resolution, and aerial photogrammetry are utilized to perform the analysis effectively. Spectroscopic analysis is performed to analyze the environmental stress level, illness and nutritional deficiency attained in the plant and also to improve the imaging approaches. Multispectral imaging is termed as the data attained from the spectral band three to seven. The multi-spectral data are utilized to analyze the quality of food, the structure of tissue, and symptoms related to plant disease. Hyper-spectral imaging is determined as the data that are attained in the range of 10 to 100 spectral bands and also they offered more data not presented in multispectral information. These hyperspectral images attained great interest in the precision farming application.

Due to the presence of a huge number of complex factors crop yield prediction is considered more challenging. The yield in a crop mainly depends on several factors such as soil quality, genotype, water accessibility, harvest planning, genotype, climatic condition, landscapes, and pest infestation. Most of the crop yield approaches are presented in non-linear nature and they are time specific. These kinds of approaches are considered highly challenging due to the presence of a huge range of interrelated factors that are affected by external aspects. Crop yield prediction is performed based on the experience of a farmer as well as historical information. In recent times, several advancements performed in validating the crop by machine learning approaches and also performed prediction more accurately. Huge numbers of predictions are done to attain effectively higher potential at the time of utilizing machine learning approaches than conventional approaches. Machine learning approaches utilized linear as well as non-linear structures to attain effective prediction rates and these kinds of approaches are attained from different learning approaches in machine learning-based agricultural systems.

5. CONCLUSION

In this work, various conventional and advanced machine learning techniques used for crop yield prediction were reviewed. It is found that machine learning based techniques can offer an effective alternative way for crop yield prediction without expert knowledge. In the recent years various advanced machine learning techniques are developed and used for crop yield prediction. Still. Some simple techniques such as linear regression are quite useful. Although, there exist a number of challenges that need to address by the researchers. Some of the potential research gaps were identified in this work to provide a potential scope for future work.

6. REFERENCES

[1] Mengjia Qiao; Xiaohui He; Xijie Cheng; Panle Li; Haotian Luo; Zhihui Tian; Hengliang Guo, "Exploiting Hierarchical Features for Crop Yield Prediction Based on 3-D Convolutional Neural Networks and Multikernel Gaussian Process," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 4476-4489, 2021.

[2] R. Luciani, G. Laneve and M. JahJah, "Agricultural Monitoring, an Automatic Procedure for Crop Mapping and Yield Estimation: The Great Rift Valley of Kenya Case," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2196-2208, July 2019.

[3] Y. Ma and Z. Zhang, "A Bayesian Domain Adversarial Neural Network for Corn Yield Prediction," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022.

[4] S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," IEEE Access, vol. 10, pp. 23625-23641, 2022.

[5] L. Martínez-Ferrer, M. Piles and G. Camps-Valls, "Crop Yield Estimation and Interpretability With Gaussian Processes," IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 12, pp. 2043-2047, Dec. 2021.

[6] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," IEEE Access, vol. 8, pp. 86886-86901, 2020.

[7] H. Aghighi, M. Azadbakht, D. Ashourloo, H. S. Shahrabi and S. Radiom, "Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 12, pp. 4563-4577, Dec. 2018.

[8] Ömer Vanli, Ishfaq Ahmad and Burak Berk Ustundag, "Area Estimation and Yield Forecasting of Wheat in Southeastern Turkey Using a Machine Learning Approach," Journal of the Indian Society of Remote Sensing vol. 48, pp. 1757–1766, 2020.

[9] H. R. Seireg, Y. M. K. Omar, F. E. A. El-Samie, A. S. El-Fishawy and A. Elmahalawy, "Ensemble Machine Learning Techniques Using Computer Simulation Data for Wild Blueberry Yield Prediction," IEEE Access, vol. 10, pp. 64671-64687, 2022.

[10] Janmejay Pant, R.P.Pant, ManojKumar Singh, Devesh Pratap Singh and Himanshu Pant,"Analysis of agricultural crop yield prediction using statistical techniques of machine learning, "Materials Today: Proceedings,Volume 46, no 20, pp. 10922-10926, 2021.

[11] Mukesh Singh Boori, KomalChoudhary, Rustam Paringer and Alexander Kupriyanov,"Machine learning for yield prediction in Fergana valley, Central Asia," Journal of the Saudi Society of Agricultural Sciences, 2 August 2022.

[12] Nguyen-Thanh Son, Chi-Farn Chen, Youg-Sin Cheng, Piero Toscano, Cheng-Ru Chen, Shu-Ling Chen, Kuo-Hsin Tseng, Chien-Hui Syu, Horng-Yuh Guo and Yi-TingZhang, "Field-scale rice yield prediction from Sentinel-2 monthly image

composites using machine learning algorithms," Ecological Informatics, Vol. 69, pp. 101618, July 2022.

[13] Minghan Cheng, Josep Penuelas, Matthew F McCabe, Clement Atzberger, Xiyun Jiao, Wenbin Wu and Xiuliang Jin, "Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China," Agricultural and Forest Meteorology, Vol. 323, pp. 109057, 15 August 2022.

[14] Moiz Uddin Ahmed and Iqbal Hussain," Prediction of Wheat Production Using Machine Learning Algorithms in northern areas of Pakistan," Telecommunications Policy, Vol. 46, no. 6, pp. 102370, July 2022.

[15] Yahui Guo, Yongshuo Fu, Fanghua Hao, Xuan Zhang, Wenxiang Wu, Xiuliang Jin, Christopher Robin Bryant and J.Senthilnath, "Integrated phenology and climate in rice yields prediction using machine learning methods," Ecological Indicators, Vol. 120, pp. 106935, January 2021.

[16] Sara Tokhi Arab, Ryozo Noguchi, Shusuke Matsushita and Tofael Ahamed, "Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach," Remote Sensing Applications: Society and Environment, Vol. 22, pp. 100485, April 2021.

[17] Juan Cao, Zhao Zhang, Fulu Tao, Liangliang Zhang, Yuchuan Luo, Jing Zhang, Jichong Han and Jun Xie, "Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches, "Agricultural and Forest Meteorology, Vol. 297, pp. 108275, 15 February 2021.

[18] Mohsen Shahhosseini,Guiping Hu,Isaiah Huber and Sotirios Archontoulis, " Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, Scientific Reports, vol. 11, no. 1606, 2021.

[19] Sushila Shidnal, Mrityunjaya V. Latte and Ayush Kapoor, Crop yield prediction: two-tiered machine learning model approach, International Journal of Information Technology, vol. 13, pp. 1983–1991, 2021.

[20] Patrick Filippi, Edward J. Jones, Niranjan S. Wimalathunge, Pallegedara D. S. N. Somarathna, Liana E. Pozza, Sabastine U. Ugbaje, Thomas G. Jephcott, Stacey E. Paterson, Brett M. Whelan and Thomas F. A. Bishop, "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning," Precision Agriculture, vol. 20, pp. 1015–1029, 2019.

[21] Dhivya Elavarasan and P. M. Durai Raj Vincent, "A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 10009–10022, 2021.

[22] Dhivya Elavarasan and P. M. Durai Raj Vincent, "Fuzzy deep learning-based crop yield prediction model for sustainable agronomical frameworks," Neural Computing and Applications, vol. 33, pp. 13205–13224, 2021.

[23] P. Murali, R. Revathy, S. Balamurali and A. S. Tayade, "Integration of RNN with GARCH refined by whale optimization algorithm for yield forecasting: a hybrid machine learning approach," Journal of Ambient Intelligence and Humanized Computing, 2020.

[24] Raí A.Schwalbert, Telmo Amado, Geomar Corassa, Luan Pierre Pott, P.V.Vara Prasadb, Ignacio A.Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," Agricultural and Forest Meteorology, vol. 284, pp. 107886, 15 April 2020.

[25] Foyez Ahmed Prodhan, Jiahua Zhang, Til Prasad Pangali Sharma, Lkhagvadorj Nanzad, Da Zhang, Ayalkibet M.Seka, Naveed Ahmed, Shaikh Shamim Hasan, Muhammad Ziaul Hoque and Hasiba Pervin Mohana," Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach," Science of The Total Environment, vol. 807, no. 3, pp. 151029, 10 February 2022.

[26] Khaki S, Wang L and Archontoulis SV (2020) A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science, vol. 10, pp. 1750, 2020.

[27] Muruganantham, Priyanga, Santoso Wibowo, Srimannarayana Grandhi, Nahidul Hoque Samrat, and Nahina Islam, "A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing" Remote Sensing, vol. 14, no. 9, pp. 1990, 2022.

[28] Ansarifar, J., Wang, L. & Archontoulis, S.V., "An interaction regression model for crop yield prediction," Scientific Reports, vol. 11, pp. 17754, 2021.

[29] Geipel, J., Link, J., Claupein, W., "Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system," Remote sensing vol. 6, pp. 10335–10355, 2014.

[30] Huang, H., Wei, X., Zhou, Y., "An overview on twin support vector regression. Neurocomputing vol. 490, pp. 80–92, 2022.

[31] Maimon, O.Z., Rokach, L., 2014. Data mining with decision trees: theory and applications. vol. 81, 2014

[32] Choudhary, K., Shi, W., Dong, Y., Paringer, R., Random forest for rice yield mapping and prediction using sentinel-2 data with google earth engine. Advances in Space Research, vol. 70, pp. 2443-2457, 2022

[33] Nevavuori, P., Narra, N., Lipping, T., "Crop yield prediction with deep convolutional neural networks," Computers and Electronics in Agriculture vol. 163, pp. 104859, 2019.