International Journal of Modern Agriculture, Volume 9, No.4, 2020
ISSN: 2305 - 7246

CROP YIELD PREDICTION IN PRECISION AGRICULTURE USING MACHINE LEARNING TECHNIQUES: A STUDY

Rakesh Kumar Ray[1], Sujata Chakravarty[2], SMIEEE

[1,2] Centurion University of Technology & Management, Odisha, India
Email: [1]rakesh.ray@cutm.ac.in, [2]sujata.chakravarty@cutm.ac.in

Abstract

In the field of agriculture, the accurate estimation of yield production is crucial for farmers and Govt. Remote sensing plays an important role which collects data from the field level, that is been deployed in the contemporary farming system for building decision-making tool, which can predict accurate yield production and other field level parameters by minimizing operation cost. The remote sensing approach generates a lot of data gathered from different platforms, so it becomes essential to introduce the Machine Learning technique to manage the huge data. Machine Learning has the capability to analyze those huge numbers of inputs and handle the non-linear task to produce knowledge, which can use in decision making. This paper discussed several research works for irrigation control using machine learning techniques, plant disease monitoring, accurate yield production. This paper concludes that the hybrid machine learning technique will give an effective and accurate model for predicting yield production by using remote sensing for decision making and environment state estimation in agriculture.

Key words: Machine Learning, Smart Irrigation, remote sensor

Introduction

India is the land of cultivation. Above 70% of people, livelihoods are depending on cultivation. As the population of India has been increasing constantly, a proportional rise in the production of food must be accomplished. Weather condition, soil quality, fertilizer management are few factors for agriculture, where farmers use their experience to sense and implement in farming to get their crop. It is inaccurate because of frequent climate change. Keeping in mind, improving crop yield production and quality by minimizing the cost of operation and environment is a crucial target in agriculture. The potential growth and yield depend on many different agricultural attributes such as soil property, weather, land size, irrigation, and fertilizer management. Sensing those factors in due time and accurately is become a necessity in agribusiness. The ground-based vehicle, satellite, aircraft, radiometer, etc are used for collecting information for sensing inputs.

Precision weed management gets timely and accurate information, by using remote sensings such as photography, video, satellite, and airborne multi-spectral scanning [1]

Water management can be improved to irrigate crops by identifying spatial dissimilarity in crop water status through thermal remote sensing via airborne thermal imagery[2].

Recent research works are going on remote sensing technique which is applied on several crop species like maize [3], trees[4], wheat [5], and rice, wheat, maize [6]. Similarly, many other works have been done on PA tasks such as weed mapping[7], evaluating crop water and nitrogen status [8], assessing evapotranspiration and drought stress [9], and mapping of soil properties [10].

Studies have been done for estimating different vegetation properties by using remote sensing at visible and near-infrared wavelengths. This focus on the chlorophylls and other photosynthetic amounts. Xue and Su [11] reviewed recently more than 100 vegetation indices along with their environment, representativeness, applicability, and implementation precision. They determined that the usage of any current vegetation indices needs vigilant attention to the strengths and weaknesses of those indices and the definite environment for real-world applications where they will be practiced. Jaafar and Ahmad [12] and many others have been tried to

make crop yield predictions by implementing remotely sensed vegetation indices. Wireless sensor network data has been implemented using different algorithms on PA. Aqeel-ur-Rehman [13] has done a review in different aspects of agriculture on wireless sensor network and their applications and described the present system in PA.

Anna Chlingaryan [14] has done a review on machine learning for crop yield prediction and nitrogen status in precision agriculture. Machine Learning techniques and sensing techniques are rapidly developing, which will give cost-effective and complete solutions for environment state estimation, crop estimation, and decision making for the agriculture industry. A hybrid combination of Machine Learning techniques and sensing techniques is the future of agriculture.

The agriculture sector is facing challenges of producing more to feed the demand for food because of the population with reducing environmental degradation, the use of natural resources, and adapting to change of climate. Smart agriculture is a solution, sensor data are collected, selected, and analyzed using ML techniques to solve various agriculture problems, accurate yield prediction, seeds selection, disease detection, fertilizers use, etc. Hassina [15] has done a review in the practice of Data Mining on smart agriculture. The paper is organizes as follows: section 2 is an overview of different machine learning techniques, section 3 is a deals with the related work on irrigation control using machine learning and sensor, section 4 is a review of papers on plant disease monitoring. In section 5, Crop Yield Estimation is given. Section 6 concludes the paper.

Machine Learning Techniques

The human brain can solve any problem what a computer can do, but it has a limitation of time to compute. Machine Learning has the advantages of computing large datasets, to pull knowledge out of those and process to solve the real-time problems very efficiently. ML empower better decision making and equipped in real-world problems can solve the problem without intervention of human. ML gives a flexible and powerful framework, which can be applied to data-driven decision making as well as expert knowledge into the system. Many domains are using Machine Learning because of these few key features of ML techniques and highly applicable in agriculture.

Machine Learning is an approach of program computers to learn from data. It is the best fit for a non-stable environment, a very complex problem where the traditional approach failed and a problem with a very long list of rules to find a solution. ML is categorized into supervised, unsupervised, and reinforced learning. In supervised learning, the data with the desired solution, are called “labels” provided to the algorithm. In unsupervised learning, data have no labels. In reinforcement learning, an Agent (AI System) observed the environment, perform assigned actions, collect the reward in return (self-learning).

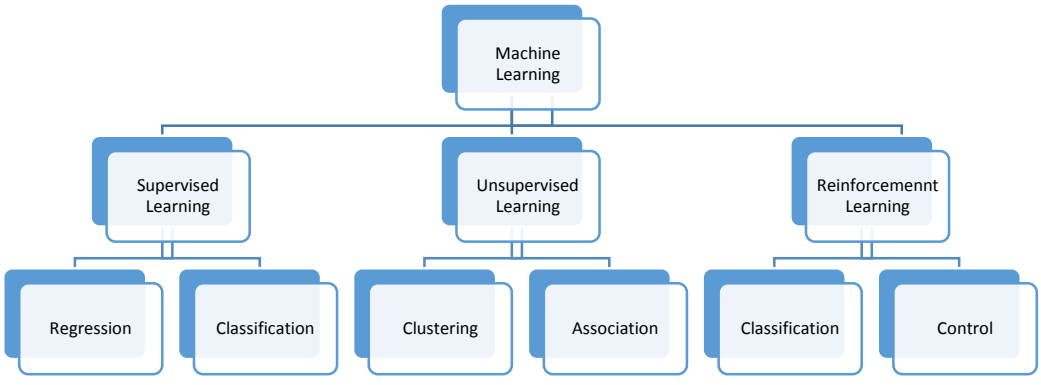

Figure 1. Classification of Machine Learning

Classification

Classification is a type of supervised learning. It predicts the class label from a set of training data and is best fitted when the label has finite and discrete values. It groups each element of data into one of a predicted class set. It has widely used in the agriculture field, like in crop disease prediction. Some of the most used classification techniques are Logistic Regression, Support Vector Machines, K-nearest Neighbors (KNN), Kernel Support Vector Machines (SVM), Naïve Bayes, Decision Tree Classification, Random Forest Classification, Neural Networks (NN), Deep Learning.

Decision Tree algorithm belongs to the family of supervised learning algorithms. decision tree algorithm can be used for regression and classification problems too. They provide a model in the form of a tree structure. Decision Trees are easy to integrate and recognize through visual presentation. The rules of classification are constructed on tests organized in the tree.

Support Vector Machine is a supervised learning algorithm proposed by Vladimir Vapnik in the year 1990. It is one of the most popular algorithms, given very good performance in many real-world applications, like image processing, bioinformatics, medical diagnostics, recently in Precision Agriculture. it can be used for regression and classification problems. SVM produces input/output mapping functions from a labeled training data set. multiclass classification is one of the possibilities.

Neural Network is a widely used supervised learning techniques in recent years. it is an attempt to recognize the closest relationship in a set of data through computation by mimics the way the human brain operates. Backpropagation is one Neural Network model.

The K-Nearest Neighbors is used for classification and regression problem form of supervised learning techniques. K is the value provided by the user, the algorithm clusters based on similarity measurement. K mean algorithm used and depending on the distance it clustered training dataset.

The Random Forest is a supervised learning algorithm, can be used for classification and regression problems. the forest builds on a combination of decision trees, usually trained with the bagging method. The vote for the most popular category decides the result of Random Forest.

Logistic Regression is a supervised Machine Learning algorithm that is used for classification problems. It is two types, Binary, and Multi-linear functions. Some applications are Email spam, online fraud detection, Tumor Malignant. The output is transformed by Logistic Regression using the logistic sigmoid function to return a probability value.

Naive Bayes classifier works on the principle of conditional probability as given by the Bayes theorem. It is a supervised machine learning algorithm. It is very simple and easy to implement. It works on less training data, even on continuous and discrete data. It is highly scalable and fast, so it can implement in real-time predictions.

Deep Learning is a sub-field of machine learning, which can be supervised, unsupervised, or semi-supervised. It uses Artificial Neural Network with hidden layers with representation learning. Deep Learning is used in many domains in recent times. In agriculture, it can be used for crop disease detection, soil quality through satellite images.

2.2 Clustering

Clustering is an unsupervised problem of finding natural groups in the feature space of input data. It divides objects into clusters(Groups) based on data, which describes objects and their relationship. Clustering operation is performed on a pattern among objects, objects with similar patterns clubbed into one cluster, likewise, there can be different clusters from a given dataset. Clustering techniques are unsupervised learning so when the data have no predefined classes and instances must group in different clusters based on their

relationships. They are classified according to Partitioning methods, density-based methods, Grid-based methods, hierarchical methods, and Model-based methods.

Partitioning Methods partition the objects into k clusters and each partition procedure one cluster. This method is implemented to optimize an objective criterion. K-Means is the popular unsupervised learning algorithm that solves the clustering problem. K-means algorithm group n observations into k clusters. The clustering operation is done with the nearest mean serving as a prototype of the cluster for each observation. It can serve the agriculture domain to detect diseases from image segmentation.

Density-Based Methods have good accuracy and the ability to merge two clusters. These methods group objects based on dense regions having some similarities. OPTICS, DENCLUE, DBSCAN [15] are the algorithms of this category.

In Grid-based Methods, the data space is framed into a finite number of cells that form a grid-like structure. All the grouping operations are done on these grids. STING and CLIQUE are some algorithms that belong to this category[15].

Hierarchical methods create a tree-type structure of clusters. Divisive (top-down approach) and agglomerative (bottom-up approach) are two divisions of this category. CHAMELEON, Balanced Iterative Reduction, and Clustering using Hierarchies (BIRCH) are some algorithms of this category [15].

2.3 Association

Association is a rule-based machine learning method. It discovers an interesting relationship between variables in a large dataset. Based on the concept of strong rules discover patterns between elements or determine association in the dataset by identifying the most common groups of data in a dataset that appear together. Apriori, AprioriTid, Dic, Eclat, and FP-growth are some algorithms used for association.

2.4 Regression

Regression is concerned with modeling the relationship between variables that is iteratively refined using a measure of the error in the predictions made by the model. It is applied with output variable is a real or continuous value, such as house price, salary. Analyzing patterns of data for predicting future values is the job of regression. It is also known as the prediction. Time series analysis and Regression Analysis are used in Regression.

A sequence of observed data with respect to time is referred to as a time series problem. Statistical modeling is used to discover hidden insides and knowledge from the time-series data in a given period of time. Predicting future values from previous observations is the aim of the forecasting model.

Regression analysis is the statistical model to find and analyze relationships between one or more variables. It predicts the dependent variables from the independent variable or variables. In this approach, the relationship among the variables can be presented with a linear model. SVM, linear regression, logistic regression, Neural network, etc. are some popular algorithms are used to forecast the future predicted values for different domains. Crop yield prediction, soil moisture prediction, pest prediction, disease prediction are some regression applications in the agriculture domain.

2.5 Fuzzy Logic

Fuzzy logic gives the degree of membership to the feature. The term fuzzy logic was introduced in 1965 by Lotfi Zadeh a professor of UC Berkeley in California. Computer logic was not capable of process data representing subjective or unclear human ideas. So fuzzy logic has been applied to those data to represent in a quantitative fashion which can be used in the computation. In the real-world, some data sets are subjective and

unclear where common reasoning patterns are present, which can not be used directly in computation. Fuzzy logic is a purely mathematical concept, adapted to handle all such subjective and unclear. It applies mathematical rules to represent human skills so data can be better understood. Fuzzy is a knowledge-based system where logical reasoning relies on rules and fuzzy sets are functions. In Machine Learning, fuzzy-based systems can be used in medical decision making, risk analysis, multimedia application, gene network analysis, economics, etc. Fuzzy based clustering algorithms can be applied to an application where objects belong to multiple clusters like similar symptoms for different diseases. Fuzzy C Means Clustering (FCM), Possibilistic C Means Clustering(PCM), Fuzzy Possibilistic C-Means Algorithm (FPCM), Possibilistic Fuzzy C Means Algorithm (PFCM) are some fuzzy-based clustering algorithms.

2.6 Genetic Algorithm

Genetic Algorithm is biological inspired techniques, used to solve complex optimization problems. The basic idea was to change the population (input) such that we get better output. GAs was developed by Prof. JohnHolland at the University of Michigan during the 1960s and 1970s. GA follows steps like initialization, fitness assignment, selection, crossover, and mutation. This follows the survival of fittest. The algorithm starts with the initialization of the population on which a classification makes on their fitness and then is used to create a new generation of the population. The process anticipates that the new generation population will be an improved one, which going to replace the old one. The selection of new improved generation based on the fitness of the population. This is a cyclic process, where each iteration population is been optimized. Applications of GA are feature selection, implementation using the Tree-based Pipeline Optimization Technique(TPOT). It can be used in Engineering Design, travel Salesman Problem, Robotics, etc. They are used to improve the performance of algorithms for example Association Rules and Decision Tree algorithm.

2.7 PCA (Principal Component Analysis)

PCA is one of the popular Machine Learning techniques. PCA is predominantly used as a dimensionality reduction technique in domains like facial recognition, computer vision, and image compression. It can be used when we want to reduce the number of variables that were unable to identify the non-considerable variables. It also can be used where variables are independent of one another. This method is insensitive to data with noise. It is a statistical procedure to produce a set of uncorrelated values from a set of correlated variables.

Irrigation control using Machine Learning and Sensor

Water is life for all living things. The use of excess or insufficient quantity of water has a negative impact on the growth of plants. Risk of salinization increase with poor water quality. Excess irrigation can be a cause of amoeba cysts, pathogens pseudomonas, etc. in a crop. Many smart irrigation systems have been developed based on Machine learning techniques in order to fulfill crop needs by considering climate and session. Machine Learning plays a vital role in ensuring quality management of irrigation with respect to consumption of water evaluation using factors like weather, crop type, and session and cost.

People face a problem watering their plants when they are on leave or at work. Tiwari [16] has proposed a mobile-based application to monitor relative humidity, atmospheric temperature, and soil moisture of plants to watering plants by start and stop a water pump using a mobile application. Sensors have been used to sense the environmental data.

An automated irrigation system has been proposed in the paper [17] to control and monitor the needs of water in the field. The sensors are installed in the field to collect and monitor different parameters like soil temperature, humidity, and soil type(Ph). The naive Bayes technique is applied for the estimation of the amount of water required for a crop in the field. Using a mobile app users can check the status of different parameters. Apart from this weather forecast has been implemented to regulate the quantity of water supplied to the field automatically and also suggests fertilizers use for crops.

For optimum and effective utilization of water in irrigation is essential. Smart irrigation system to be developed on dynamic data collected from the field. A paper [18] presented a working model to predicts soil moisture based on collected sensors data deployed at the field and forecasted weather data available on the internet. k-means clustering and Support Vector Regression (SVR) used for estimation of change in soil moisture due to weather conditions.

The paper [19] has proposed a model for estimation of vineyard water status using two machine learning techniques (Rotation Forests and Decision Trees) and on-the-going thermal images which are potential for decision making on irrigation.

The paper [20] has proposed a real-time application to predict weekly irrigation recommendation plan along with monitoring and controlling crop. 22 sensors are deployed in four major plots (which are divided into eight irrigation groups and 28 subplots) to gather data. for regression model Gradient Boosted Regression Trees used which has given 93% accuracy and classification model Boosted Tree classification used which gives 95% accuracy. The model can significantly facilitate the irrigation planning process.

The paper [21] has proposed an on-demand irrigation system by using numerical weather prediction(NWP) and Time-of-use(TOU). It estimates the amount of energy and time required to perform the next operation. The Support Vector Regression method is been used to forecast solar energy. On-demand, scheduling reduce the irrigation cost by optimization. The result showed that water resources and energy can be held back by 7.97% and the amortized price can be decreased by 25.34% while comparing with the irrigation method built on soil moisture.

The paper [22] has proposed a neuro-fuzzy computational algorithm to manage irrigation called FITRA. Sensors are been deployed in the field to collect field data periodically, stored at the central cloud system. The model controls the irrigation process based on sensor data. The system is adaptive in nature, which changes automatically with environmental conditions. The model can increase production and minimize consumption.

This paper [23] has proposed a model for an automatic irrigation system to save water. The Fuzzy-Logic-based controller monitors the tank water level and, the rain amount, the speed of the wind, and the atmospheric temperature. The sensor nodes collect temperature and moisture, the coordinator node, the irrigation controller using Fuzzy Logic to decide the irrigation timing and to apply to water, the irrigation pipe network are parts of the model. The system may be sent information to the farmer if necessary through the GSM module.

Plant Disease Monitoring

Plant growth and productivity are affected by several diseases. The earliest detection is the only way to overcome the situation. Machine learning can help to overcome using image processing techniques. In recent years, several research is giving remarkable results on detecting early weed, insect pests in the crop, and plant diseases [24]. Many Machine learning techniques have been implemented for precision agriculture such as neural networks, support vector machines for classification and self-organized maps, k-means for clustering.

The paper [25] has proposed a machine learning application to diagnose crop disease on plant images. The disease severity and disease incidence are important factors. The disease severity has 5 classes and disease incidence has 4 classes. LinearSVC, KNN, and Extremely Randomized Trees algorithms are implemented for classification where 99% accuracy is being scored in the classification of disease severity. An android-based application developed to implement the model where a farmer can take photos from a smartphone and can get the result.

Production of crop suffers because of crop diseases. This paper [26] has proposed a model to detect crop disease. deep convolutional neural network algorithm is being used for the purpose, where the model achieved 99.35% accuracy. the experiment was conducted on a public dataset on 14c crops and 26 diseases. the smartphone can be used for the implementation of the model, as the users are increasing constantly.

The paper [27] has proposed an intelligent model to detect the plant stresses or diseases on multi-sensors. The model used the Geostationary Positioning System (GPS), hyperspectral and multi-spectral information. The model used two classifiers: Quadratic Discrimination (QDA) and Neural Networks. To avoid fungal diseases, a site-specific spraying approach is proposed.

The paper [28] has proposed a model to detect the Maize Streak Virus (MSV) in South Africa. RapidEye satellite data processed by Random Forest classifier. The field data and remote sensing data are also collected to supplement the work. The integration of vegetation indices in classification improves accuracy in the result. The red index, soil adjusted vegetation index, enhanced vegetation index, normalized vegetation index are the most important in detecting MSV.

The paper [29] has used Support Vector Machine to predict wheat rust in China. In the paper, SVM is been compared with other Registration analysis methods and the SVM method has given better accuracy of fit and prediction as a result. The prediction used three datasets (disease and methodological data). The disease prediction of the application is feasible and effective.

The paper [30] has proposed a study on culture-weather-disease interactions using different Machine Learning techniques like Naive Bayes classification with Gaussian distribution, Multi-Variate Regression extraction (MVR), and Rapid Association Rule extraction. A wireless sensor network is used for leaf spot disease (LS) to collect groundnut crop monitoring data. the proposed model has been compared and validated to Agriculture Research Institute (ARI) model.

The paper [31] has proposed CNN models (AlexNet, GoogleNet) to classify tomato diseases which are nine types. They used a large dataset of tomato leaves infected diseases contains 14828 images. To understand symptoms and analysis proposed deep models, they have used visualization methods. The obtained result has shown that CNN models are better than SVM and RF (with or without pre-training).

Kawasaki [32] developed a plant disease detection system using CNN (convolutional neural networks). The developed model is detecting cucumber leaves diseases (Zucchini mosaic virus and melon yellow spot virus). To trained the model, 800 cucumber leaf images are used. The model achieved an accuracy of 94.9% under the 4-fold cross-validation strategy.

The paper [33] has proposed a Deep-CNN (convolutional neural networks) based model to detect five apple leaf diseases in real-time. 26,377 images are used for testing the model. The proposed INAR-SSD model has shown the performance of 78.80% mAP on the apple leaf disease dataset, with a detection speed of 23.13 FPS.

This paper has proposed a model [34] using Wireless Sensor Network to process real-time using machine learning algorithms to predict plant diseases and pests. The model has developed using Navie Bayes Kernel and WSN. The proposed model has used Navie Bayes Kernel to predict plant diseases and pests. The crop dataset consists of Crop name, Soil values (Nitrogen, Potassium, Phosphor), pH values, Humidity, Temperature, and Pest dataset consists of Pest name, Crop name, humidity, the temperature used to predict the pattern of plant diseases and pest. Wireless sensor networks are used to get data from the field.

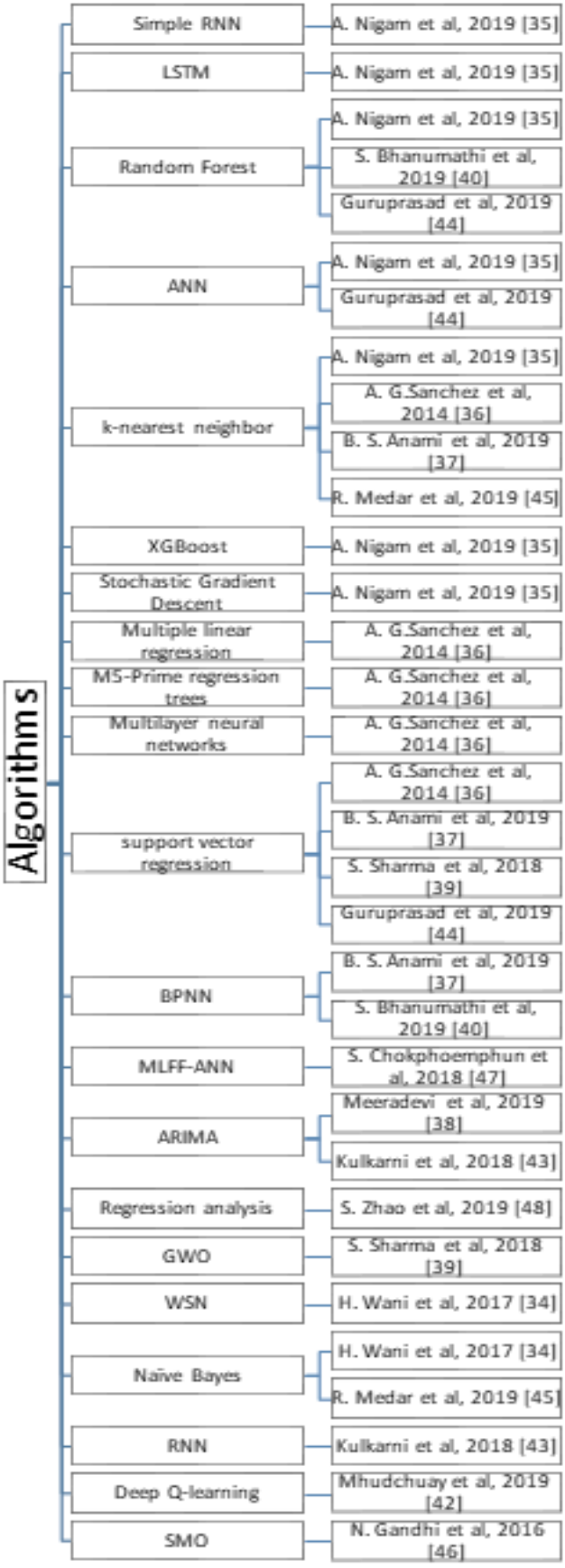
Algorithms
Simple RNN
A. Nigam et al, 2019 [35]
LSTM
A. Nigam et al, 2019 [35]
Random Forest
A. Nigam et al, 2019 [35]
S. Bhanumathi et al, 2019 [40]
Guruprasad et al, 2019 [44]
ANN
A. Nigam et al, 2019 [35]
Guruprasad et al, 2019 [44]
k-nearest neighbor
A. Nigam et al, 2019 [35]
A. G.Sanchez et al, 2014 [36]
B. S. Anami et al, 2019 [37]
R. Medar et al, 2019 [45]
XGBoost
A. Nigam et al, 2019 [35]
Stochastic Gradient Descent
A. Nigam et al, 2019 [35]
Multiple linear regression
A. G.Sanchez et al, 2014 [36]
M5-Prime regression trees
A. G.Sanchez et al, 2014 [36]
Multilayer neural networks
A. G.Sanchez et al, 2014 [36]
support vector regression
A. G.Sanchez et al, 2014 [36]
B. S. Anami et al, 2019 [37]
S. Sharma et al, 2018 [39]
Guruprasad et al, 2019 [44]
BPNN
B. S. Anami et al, 2019 [37]
S. Bhanumathi et al, 2019 [40]
MLFF-ANN
S. Chokphoemphun et al, 2018 [47]
ARIMA
Meeradevi et al, 2019 [38]
Kulkarni et al, 2018 [43]
Regression analysis
S. Zhao et al, 2019 [48]
GWO
S. Sharma et al, 2018 [39]
WSN
H. Wani et al, 2017 [34]
Naïve Bayes
H. Wani et al, 2017 [34]
R. Medar et al, 2019 [45]
RNN
Kulkarni et al, 2018 [43]
Deep Q-learning
Mhudchuay et al, 2019 [42]
SMO
N. Gandhi et al, 2016 [46]

Figure 2. Machine Learning Algorithms used in Crop Yield Estimation

Crop Yield Estimation

In agriculture, farmers need to make decisions every day based on factors like the variability of soil, environmental conditions, commodity price, etc. J. Majumdar et al[35] applied data mining techniques like CLARA, PAM, DBSCAN, and Multi Linear Regression to maximize crop production using optimal parameters founded by analyzing the agriculture data.

Crop yield prediction matter on some factors like rainfall, temperature, season, and area. Nigam [36] presented a study of these machine learning techniques (Simple RNN, LSTM, Random Forest, ANN, KNN, XGBoost, Stochastic Gradient Descent) for crop yield prediction. LSTM predicts better in case of Temperature and Sample RNN predict better in case of Rainfall and random forest classifier gives better accuracy for yield prediction. The result of the test ranked best to worst are Random Forest, XGBoost, KNN, Logistic Regression.

In another study, Sanchez [37] has applied these ML techniques, Multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression, and k-nearest neighbor methods to predict crop yield production on 10 numbers of the dataset. 8 numbers of attributes (Area, water, solar, rainfall, maxTemp, MinTemp, AvgTemp, Season-duration) has been used in the reached work. M5-prime regression gives the best result among all other techniques. For validation, mean absolute error (MAE), root mean square error (RMS), root relative square error (RRSE), and correlation factor (R) are been used.

Back Prorogation Neural Network gives the best result in Stress Management System, to detect biotic and abiotic paddy stress. The paper has [38] using machine learning techniques like BPNN, SVM, and KNN on 2D images using the state of art color feature. For feature extraction used sequential forward floating selection. In his work, they Annalise on 2 bionic and 9 abiotic paddy stress.

In a study, Meeradevi [39] has applied ARIMA, linear regression methods to predict yield production and select crop for cultivation, on the Govt of India dataset. The model has been developed using Android App for farmers to suggest the crop and quantities of production depending on factors like rainfall, farm size, temperature, location. ARIMA method has been used for forecasting rainfall and temperature. Linear regression is used to predict the yield production and select crop for cultivation, depending upon the data provided by farmers.

Sharma [40] has come up with a model that predicts crop using a hybrid Machine learning algorithm, which will tell whether a bad or good yield using Hadoop, SVM, GWO. The proposed work used food Agriculture Organization data. To process the data Hadoop has been used. For feature extraction, Gray Wolf Optimization has been used. SVM classifier has been used for prediction. The proposed model has not given high accuracy.

This paper [41] has have proposed a model using Machine Learning techniques to predict yield production as well as the amount of fertilizer to be used to get maximum production to help farmers in India depending on the types of soil. Random Forest and Back-Propagation algorithms are used in the proposed model. The proposed work has used 7 factors as state, district, crop, area, season, and production to predict yield production using random forest. Fertilizer is used for increasing crop productivity, but the right amount and time is crucial. A back-propagation algorithm has been used to predict the amount of fertilizer to be used by farmers depending on soil parameters as nitrogen, phosphorus to get maximum yield.

This paper [42] has proposed a model called CSM (Crop Selection Model), which helps to increase the yield rate of crops over a season. The work introduced a crop sowing table with fields as crop name, sowing period, harvesting period, growing days, predicted yield rate. CSM has been developed on the table which gives a sequence of crops that can increase the yield ratio of a season depending on the input parameter.

This paper [43] has have proposed a working model on rice crop cultivation practice to suggest farmers when to start cultivation and when to harvest crop so farmers can get maximum productivity using Deep Q-learning. The paper proposed a model using Deep Q learning neural network to when to cultivate and harvest. The historical climate data has been used to model so farmers can get maximum profit from cultivating rice as a crop. The paper has limited its work on climate data only, which can be further enhanced to when to fertilize and use pesticide.

This paper [44] has proposed a model for predicting rice yield production (tone per hector) using a data-driven hybrid machine learning model, on the district of Karnataka, which can helpful for farmers to take necessary action. The model used ARIMA and RNN machine learning algorithms. The features used for building the model were Rainfall, Soil properties (Nitrogen, phosphorous, potassium, scale of pH of soil). The rainfall has been predicted using the time series algorithm ARIMA. The proposed work gives a high error rate, which can be optimized.

This paper [45] has proposed a model on the yield prediction of paddy on soil and weather parameters using machine learning technique as Random Forest, Neural Network, Support Vector Machine (3 kernels) on historical data of Indian Government. The work has been implemented on the Indian Government dataset. The dataset is broken down to taluk. 3 machine learning technique has experimented weekly, so the closest and accurate prediction can be obtained.

This paper [46] has proposed a working application model using Naïve Bayes and KNN for predicting yield rate and selection of crop. KNN and Naïve Bayes are used for predicting yield rate and crop selection, which will helpful for the farmer. The author has used Weka tool to develop the model. It has been observed that both machine learning techniques can be used. Crop name, season, and district are used as attributes for the development of the model.

The paper [47]is an experimental study on rice yield prediction by implementing SMO classifier. The experiment shows that SMO is performing the lowest accuracy and low performance as compared to BayesNet and Multilayer perceptron. WEKA is a tool used for performing operations. The Govt of India dataset is been used on Maharashtra state. The Author used parameters as Minimum, maximum & average temperature, area, precipitation, Evapotranspiration, production,

The paper[48], The Machine learning technique Multi-Layer Feed Forward ANN model used for predicting moisture ratio at the time of drying process to control environment factor to increase productivity in agriculture. The regression coefficient of determination (R2) of 0.99556 mean absolute error (MAR) of 0.00127 and mean squared error of 1.988 * 10-4. This is a simulation-based application for predict paddy drying behaviorists improved by applying the swirling flow principle.

This paper [49] has proposed a model on 2D images to predict rice yield measurement on rice panicles. In the work panicle weight parameter, the panicle grain area assumed height process through regression analysis which obtains accuracy up to 95%. In the model regression analysis is used for determinants of grain weight. Grain area and weight parameter correlation were calculated and verified for different cultivars.

Conclusion

Machine Learning and Sensing Technology are rapidly improving during the last decade. The development is also constantly happening with more sophisticated algorithms and hybrid models. These technologies are showing new and advanced dimensions, which can be applied to the datasets of different crops and the best and appropriate solutions can be attained for the existing problems in agriculture. This paper discussed several efforts and solutions towards the agriculture domain.

References

[1] D. W. Lamb and R. B. Brown, "Remote-sensing and mapping of weeds in crops," *J. Agric. Eng. Res.*, vol. 78, no. 2, pp. 117–125, 2001, doi: 10.1006/jaer.2000.0630.
[2] A. K. Tilling *et al.*, "Remote sensing of nitrogen and water stress in wheat," vol. 104, pp. 77–85, 2007, doi: 10.1016/j.fcr.2007.03.023.
[3] S. Taghvaeian, J. L. Chávez, and N. C. Hansen, "Infrared thermometry to estimate crop water stress index and water use of irrigated maize in northeastern colorado," *Remote Sens.*, vol. 4, no. 11, pp. 3619–3637, 2012, doi: 10.3390/rs4113619.
[4] J. Bellvert *et al.*, "Airborne thermal imagery to detect the seasonal evolution of crop water status in peach, nectarine and Saturn peach orchards," *Remote Sens.*, vol. 8, no. 1, pp. 1–17, 2016, doi: 10.3390/rs8010039.
[5] N. K. Gontia and K. N. Tiwari, "Development of crop water stress index of wheat crop for scheduling irrigation using infrared thermometry," *Agric. Water Manag.*, vol. 95, no. 10, pp. 1144–1152, 2008, doi: 10.1016/j.agwat.2008.04.017.
[6] W. G. M. Bastiaanssen and P. Steduto, "The water productivity score (WPS) at global and regional level: Methodology and first results from remote sensing measurements of wheat, rice and maize," *Sci. Total Environ.*, vol. 575, pp. 595–611, 2017, doi: 10.1016/j.scitotenv.2016.09.032.
[7] R. Sui, J. A. Thomasson, J. Hanks, and J. Wooten, "Ground-based sensing system for weed mapping in cotton," *Comput. Electron. Agric.*, vol. 60, no. 1, pp. 31–38, 2008, doi: 10.1016/j.compag.2007.06.002.
[8] D. M. El-Shikha, P. Waller, D. Hunsaker, T. Clarke, and E. Barnes, "Ground-based remote sensing for assessing water and nitrogen status of broccoli," *Agric. Water Manag.*, vol. 92, no. 3, pp. 183–193, 2007, doi: 10.1016/j.agwat.2007.05.020.
[9] M. Greco, A. Chiappetta, L. Bruno, and M. B. Bitonti, "In Posidonia oceanica cadmium induces changes in DNA methylation and chromatin patterning," *J. Exp. Bot.*, vol. 63, no. 2, pp. 695–709, 2012, doi: 10.1093/jxb/err313.
[10] B. & W. C. Barnes, Edward M., Sudduth, Kenneth A., Hummel, John W., Lesch, Scott M.; Corwin, Dennis L., Yang, Chenghai, Daughtry, Craig S.T., "Remote- and Ground-Based Sensor Techniques to Map Soil Properties," *Am. Soc. Photogramm. Remote Sens.*, pp. 619–630, 2003, doi: https://doi.org/10.14358/PERS.69.6.619.
[11] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, 2017, doi: 10.1155/2017/1353691.
[12] H. H. Jaafar and F. A. Ahmad, "Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands," *Int. J. Remote Sens.*, vol. 36, no. 18, pp. 4570–4589, 2015, doi: 10.1080/01431161.2015.1084434.
[13] Aqeel-Ur-Rehman, A. Z. Abbasi, N. Islam, and Z. A. Shaikh, "A review of wireless sensors and networks' applications in agriculture," *Comput. Stand. Interfaces*, vol. 36, no. 2, pp. 263–270, 2014, doi: 10.1016/j.csi.2011.03.004.
[14] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agric.*, vol. 151, no. June, pp. 61–69, 2018, doi: 10.1016/j.compag.2018.05.012.
[15] H. Ait Issad, R. Aoudjit, and J. J. P. C. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," *Eng. Agric. Environ. Food*, vol. 12, no. 4, pp. 511–525, 2019, doi: 10.1016/j.eaef.2019.11.003.
[16] R. Tiwari, Y. Patel, and G. Saha, "User Controlled Precision Irrigation System," *2019 Innov. Power Adv. Comput. Technol. i-PACT 2019*, no. October 2018, pp. 1–7, 2019, doi: 10.1109/i-PACT44901.2019.8959516.
[17] P. Padalalu, S. Mahajan, K. Dabir, S. Mitkar, and D. Javale, "Smart water dripping system for agriculture/farming," *2017 2nd Int. Conf. Converg. Technol. I2CT 2017*, vol. 2017-Janua, pp. 659–662, 2017, doi: 10.1109/I2CT.2017.8226212.
[18] A. Goap, D. Sharma, A. K. Shukla, and C. Rama Krishna, "An IoT based smart irrigation management system using Machine learning and open source technologies," *Comput. Electron. Agric.*, vol. 155, no. September, pp. 41–49, 2018, doi: 10.1016/j.compag.2018.09.040.
[19] S. Gutiérrez, M. P. Diago, J. Fernández-Novales, and J. Tardaguila, "Vineyard water status assessment using on-the-go thermal imaging and machine learning," *PLoS One*, vol. 13, no. 2, pp. 1–18, 2018, doi: 10.1371/journal.pone.0192037.
[20] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge," *Precis. Agric.*, vol. 19, no. 3, pp. 421–444, 2018, doi: 10.1007/s11119-017-9527-4.
[21] T. Xie, Z. Huang, Z. Chi, and T. Zhu, "Minimizing amortized cost of the on-demand irrigation system in

smart farms," *Proc. - 2017 3rd Int. Work. Cyber-Physical Syst. Smart Water Networks, CySWATER 2017*, pp. 43–46, 2017, doi: 10.1145/3055366.3055370.

[22] G. Kokkonis, S. Kontogiannis, and D. Tomtsis, "FITRA -A Neuro-fuzzy computational algorithm approach based on an embedded water planting system," *ACM Int. Conf. Proceeding Ser.*, pp. 0–7, 2017, doi: 10.1145/3018896.3018934.

[23] J. R. P. Anand, J., Perinbam, "Automatic irrigation system using fuzzy logic," *AEIJMR*, pp. 1–9, 2014.

[24] J. Behmann, A. K. Mahlein, T. Rumpf, C. Römer, and L. Plümer, "A review of advanced machine learning methods for the detection of biotic stress in precision crop protection," *Precis. Agric.*, vol. 16, no. 3, pp. 239–260, 2015, doi: 10.1007/s11119-014-9372-7.

[25] G. Owomugisha and E. Mwebaze, "Machine learning for plant disease incidence and severity measurements from leaf images," *Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016*, pp. 158–163, 2017, doi: 10.1109/ICMLA.2016.126.

[26] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Front. Plant Sci.*, vol. 7, no. September, pp. 1–10, 2016, doi: 10.3389/fpls.2016.01419.

[27] D. Moshou *et al.*, "Intelligent multi-sensor system for the detection and treatment of fungal diseases in arable crops," *Biosyst. Eng.*, vol. 108, no. 4, pp. 311–321, 2011, doi: 10.1016/j.biosystemseng.2011.01.003.

[28] I. Dhau, E. Adam, K. K. Ayisi, and O. Mutanga, "Detection and mapping of maize streak virus using RapidEye satellite imagery," *Geocarto Int.*, vol. 34, no. 8, pp. 856–866, 2019, doi: 10.1080/10106049.2018.1450448.

[29] H. Wang and Z. Ma, "Prediction of wheat stripe rust based on support vector machine," *Proc. - 2011 7th Int. Conf. Nat. Comput. ICNC 2011*, vol. 1, pp. 378–382, 2011, doi: 10.1109/ICNC.2011.6022095.

[30] A. K. Tripathy *et al.*, "Knowledge discovery and Leaf Spot dynamics of groundnut crop through wireless sensor network and data mining techniques," *Comput. Electron. Agric.*, vol. 107, pp. 104–114, 2014, doi: 10.1016/j.compag.2014.05.009.

[31] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315, 2017, doi: 10.1080/08839514.2017.1315516.

[32] Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic Study of Automated Diagnosis of Viral Plant Diseases Using Convolutional Neural Networks," *Springer Verlag*, vol. 9475, no. 1, pp. 842–850, 2015, doi: 10.1007/978-3-319-27863-6.

[33] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.

[34] H. Wani and N. Ashtankar, "An appropriate model predicting pest/diseases of crops using machine learning algorithms," *2017 4th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2017*, pp. 4–7, 2017, doi: 10.1109/ICACCS.2017.8014714.

[35] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *J. Big Data*, no. 20, 2017, doi: https://doi.org/10.1186/s40537-017-0077-4.

[36] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," *Proc. IEEE Int. Conf. Image Inf. Process.*, vol. 2019-Novem, pp. 125–130, 2019, doi: 10.1109/ICIIP47207.2019.8985951.

[37] A. Gonzalez-Sanchez, J. Frausto-Solis, and W. Ojeda-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction," *Spanish J. Agric. Res.*, vol. 12, no. 2, pp. 313–328, 2014, doi: 10.5424/sjar/2014122-4439.

[38] B. S. Anami, N. N. Malvade, and S. Palaiah, "Classification of yield affecting biotic and abiotic paddy crop stresses using field images," *Inf. Process. Agric.*, vol. 7, no. 2, pp. 272–285, 2020, doi: 10.1016/j.inpa.2019.08.005.

[39] Meeradevi and H. Salpekar, "Design and Implementation of Mobile Application for Crop Yield Prediction using Machine Learning," *2019 Glob. Conf. Adv. Technol. GCAT 2019*, pp. 1–6, 2019, doi: 10.1109/GCAT47503.2019.8978315.

[40] S. Sharma, G. Rathee, and H. Saini, "Big data analytics for crop prediction mode using optimization technique," *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 760–764, 2018, doi: 10.1109/PDGC.2018.8746001.

[41] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop yield prediction and efficient use of fertilizers," *Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019*, pp. 769–773, 2019, doi: 10.1109/ICCSP.2019.8698087.

[42] R. Kumar, M. P. Singh, P. Kumar, and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," *ICSTM*, no. May, pp. 138–145, 2015, doi: 10.1109/ICSTM.2015.7225403.

[43] T. Mhudchuay, T. Kasetkasem, W. Attavanich, I. Kumazawa, and T. Chanwimaluang, "Rice cultivation planning using a deep learning neural network," *Proc. 16th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2019*, pp. 822–825, 2019, doi: 10.1109/ECTI-CON47248.2019.8955227.

[44] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada, and Meeradevi, "Predictive Analysis to Improve Crop Yield using a Neural Network Model," *2018 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2018*, pp. 74–79, 2018, doi: 10.1109/ICACCI.2018.8554851.

[45] R. B. Guruprasad, K. Saurav, and S. Randhawa, "Machine Learning Methodologies for Paddy Yield Estimation in India: a Case Study," *IGARSS 2019 - 2019 IEEE Int. Geosci. Remote Sens. Symp.*, pp. 7254–7257, 2019, doi: 10.1109/igarss.2019.8900339.

[46] R. Medar, V. S. Rajpurohit, and Shweta, "Crop Yield Prediction Using Machine Learning Algorithms," *IEEE 5th Int. Conf. Converg. Technol.*, vol. Mar 29-31, pp. 1–5, 2019, doi: 10.1109/I2CT45611.2019.9033611.

[47] N. Gandhi, O. Petkar, L. J. Armstrong, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," *2016 13th Int. Jt. Conf. Comput. Sci. Softw. Eng. JCSSE 2016*, no. 2010, pp. 11–15, 2016, doi: 10.1109/JCSSE.2016.7748856.

[48] S. Chokphoemphun and S. Chokphoemphun, "Moisture content prediction of paddy drying in a fluidized-bed drier with a vortex flow generator using an artificial neural network," *Appl. Therm. Eng.*, vol. 145, pp. 630–636, 2018, doi: 10.1016/j.applthermaleng.2018.09.087.

[49] S. Zhao, H. Zheng, M. Chi, X. Chai, and Y. Liu, "Rapid yield prediction in paddy fields based on 2D image modelling of rice panicles," *Comput. Electron. Agric.*, vol. 162, no. May, pp. 759–766, 2019, doi: 10.1016/j.compag.2019.05.020.

Rakesh Kumar Ray is pursuing his Ph.D. in Centurion University of Technology and Management. His research area includes Machine Learning and smart in the field of Agriculture.

Dr. Sujata Chakravarty is a Senior Member of IEEE. Currently she is working as Professor and HoD, Dept. of CSE, Centurion University of Technology & Management, Bhubaneswar, India. She is a reviewer of many International journals like Elsevier (Neurocomputing, Knowledge-Based Systems, Econometrics and Statistics, Karbala International Journal of Modern Science), Inderscience, International Journal of Intelligent and Fuzzy Systems and IEEE. She has published 4 book-chapters and about 90 **Authors**
articles in many International journals and conferences. Her research area includes multidisciplinary fields like Application of Computational Intelligence and Evolutionary Computing Techniques in the field of Financial Engineering, Bio-medical data classification, Smart Agriculture, Intrusion Detection System in Computer-Network, Analysis and prediction of different financial time series data,