PREDICTION AND ANALYSIS OF SOIL MOISTURE USING MACHINE LEARNING TECHNIQUES

ARUN JAYAKAR S [1], SRINIVASAN S [2], MONISH G [3],
KIRUBHAKARAN K [4]
[1] Assistant Professor Level III, [1][2][3] UG scholar,
BANNARI AMMAN INSTITUTE OF TECHNOLOGY, SATHYAMANGALAM, ERODE

ABSTRACT

As the world population grows, health and adequate crop and food production are important for everyone. Crop production positively affects the country's economy. Predicting the moisture content present in the soil will help in choosing better crop for cultivation. Many advances have been made in recent years, from product selection to harvest. Selecting the right crop for cultivation increases the yield. For maximum yield, the understanding of macronutrient requirement is important. Different crops require different amount of NPK values. Temperature, humidity, rainfall, and pH are the factors need to be considered. We tend to train, test, and validate various machine learning models. Decision Tree, AdaBoost Classifier, XGB Classifier, Random Forest, Logistic Regression, SVM (Support Vector Machine) Classifier, and KNN Algorithm can be used for predicting purposes. By comparing the accuracy of various models, we can choose the efficient model. The proposed methodology is to design a user-friendly website. The user can provide inputs such as NPK values, temperature, humidity, rainfall, and pH. After the analysis by the machine learning model, the website will display the moisture content and suitable crop that can be grown in a particular region. On training, testing and validation we got higher accuracies in Logistic regression, SVM Classifier and KNN Algorithm. The accuracy of KNN algorithm came out to be 0.9886. For moisture analysis we have trained and tested using Neural Network pattern Recognition in MATLAB. By combining these two processes, we have designed a user-friendly website.

Keywords: *Macro nutrients, Soil moisture, Crop prediction, KNN, SVM, Logistic Regression.*

1. INTRODUCTION:

To be a successful farmer, one must first know the nature of the soil. Health begins from the soil. In India, to meet growing needs, it is important to make agriculture sustainable. To fulfill the needs of 1.3 billion people, more than 60% of land in the country is used for agriculture. The economy of any country depends upon the agricultural sector. In agriculture, factors such as analysis of soil, weather conditions, availability of water, and crop demand are important. The lack of knowledge of climatic variations is the main challenge faced in agriculture. The minerals present in the soil can be determined through soil testing in the lab or by using sensors. Real-time data can be collected using various sensors. Nowadays, a lot of advancements have been made, from sowing seeds to harvesting the crop. IoT, Cloud computing, and Machine learning can help cultivators in each stage of cultivation. The stages involved in agriculture can be broadly classified as choosing suitable seeds, monitoring crop growth, protecting from diseases, finding the ripening stage of the crop, and crop harvesting. A support system should be designed in such a way that it should help the farmers to take decisive measures so that farmers need not completely depend on the local agricultural offices. Macro-nutrients and micro-nutrients present in the soil play an important role in the yield of crops. Several aspects of the soil need to be analyzed before selecting a particular crop for cultivation. The growth of the plant is dependent on Nitrogen(N), Phosphorus(P), and Potassium(K) ratio in the soil. Soil properties, weather conditions, water availability, the temperature of the soil, atmospheric humidity, sunlight, wind, pollution level, etc. are factors influencing the growth of any crop. The mineral availability and soil characteristics vary from time to time. Crop yield decreases because farmers may use the wrong proportion of fertilizer without knowing actual NPK values. Soil pH and NPK relation is crucial. Machine learning (ML) can be used for pattern recognition. Mathematical (or) statistical models are used by machine learning algorithms for analysis and prediction. Machine learning is a subset of artificial intelligence that uses historical data as input to predict new output values. ML predicts based on the relationship between dependent variables and independent variables. Nitrogen, Phosphorus, and Potassium are dependent variables as they depend on soil pH, moisture, temperature, and cation exchange capacity, which are independent variables. Machine learning provides better output when there are multiple independent variables. Prediction of demanded crops can be done using machine learning with historical weather patterns and market data. Machine learning is broadly classified as supervised learning, unsupervised learning, and reinforcement learning. The dataset is divided in the ratio of 8:2, i.e., 80% of

the data is used for training, and the remaining 20% of the data is used for testing. In supervised learning, sample-labeled data is provided to the machine learning system while training. Supervised learning is further classified into classification and regression. Most profitable crops can be identified based on market demand and environmental factors. The portion of the crop above the ground is called as shoot system and the portion which is below the ground is called the root system of the plant. Shoot consists of stems, leaves, buds, flowers, and fruits. Nitrogen helps in the growth of leaves. Flowers help in pollination. Pollination is the initial process in fruit formation. The growth of roots, and flowers and the development of fruit are taken care by phosphorus. The development of root, fruit, and flower are contributed by phosphorus ions. Potassium helps in the overall functionalities of the plant. Before the 20th century, there was a lack of basic understanding of soil nutrients. In the early 20th century, research scientists discovered the Nitrogen fertilizer. For a better yield, the NPK ratio for a specific crop needs to be known. Each crop requires a different NPK ratio. Most often trendiest crop in the region is chosen by the farmer for cultivation. The fertility of lands is affected because of no rotation of crops. Selection of the wrong crops leads to changes in the pH levels of the soil. The atmospheric CO2 levels also play an important role in the health of plants as CO2 is a major component in photosynthesis.

The relationship between crop temperature and crop health is an important aspect of agricultural science. Understanding this relationship can provide valuable insights into the physiological status of crops and help farmers make informed decisions about irrigation, pest management, and overall crop management strategies. Temperature is an important environmental factor that directly affects the growth and development of crops. All crops have optimal temperatures for various physiological processes such as germination, photosynthesis, flowering, and fruiting. Higher temperatures often encourage rapid growth, but it is tedious when reaching a point where extreme heat can damage healthy crops. Fertilizer formulas are usually written with a group of three numbers representing the weight percentage of nitrogen (N), phosphorus (P2O5) and potassium (K2O) present in the fertilizer. product. For example, a commonly used fertilizer may be labeled 10-20-10; This means it contains 10% nitrogen, 20% phosphorus (as P2O5), and 10% potassium (as K2O) by weight. The proposed methodology is to predict suitable crop using an efficient machine learning algorithm by analyzing accuracy of various algorithms. By using Neural Network Pattern recognition in MATLAB, soil moisture content can be predicted. Based on these prediction and analysis we tend to develop a website.

2. LITERATURE SURVEY:

Angu raj et al., developed a system using IoT (Internet of Things) and machine learning. Soil parameters are collected by sensors and fed into the training model. The machine learning models used are Naive Bayes and Random Forest. Naive Bayes classifies objects based on their probability and thus conforms to the required classification. Naive Bayes can be used for binary distributions as well as many other types of distributions. Bayes theorem is based on the principle of Bayes rule. Bayes' law depends on probability. The results of Bayes' rule can also be explained as posterior probability. The posterior probability can be defined as the ratio of the probability multiplied by the prior probability to the relevant probability.

Murali krishna senapathy et.al., proposed an IoT enabled soil nutrient classification and crop recommendation model. In their model, they recorded humidity, temperature, NPK and pH using IoT sensors. They used the Fruit Fly Optimization (FFO) algorithm to optimize model selection for multivariate support vector machine (MVSM) models. Used for soil and water samples LDR color sensor, soil moisture sensor, pH sensor, thermometer, hydrometer, and GPS. Models such as linear SVM, kernel SVM and decision trees are used. The combination of MVSM and FFO improves classification accuracy.

Ersin elbasi et.al., in their paper explored and discussed various classification algorithms. The objective of the study is to incorporate IoT sensors and ML algorithms for waste reduction and reap production optimization. The performance of machine learning algorithms such as SVM, Naïve Bayes, Decision tree, and Random Forest were analyzed. The study identified and presented an experimental result of the opportunities and challenges in integration of IoT Sensors and ML models. Appropriate feature selection becomes an important factor to achieve better accuracy. Temperature, humidity, pH, precipitation, etc. factors serve as criteria. Using Naive Bayes, the accuracy is 97.05%. The accuracy of random forest is 97.32%.

Umm E Farwa et al. It analyzes the relationship between independent variables and variables in machine learning algorithms. Nitrogen, phosphorus, and potassium were chosen as variables. Moisture content, soil pH, cation exchange capacity CEC and temperature were chosen as independent variables. They received two records and compared the facts with these records. The relationship between NH4, NO3 and humidity temperature was analyzed. Choose from Linear Regression, Ridge Regression, Bayesian Regression and Bayesian Ridge Regression. The relationship between a variable and one or more independent variables can be obtained by horizontal lines. It will become more horizontal if more arguments are used. Multicollinearity can be eliminated using ridge regression. Ridge regression should be used when linear regression is close to optimal. Bayesian regression provides a good example of regression models. Bayesian regression is flexible because it uses data.

3. OBJECTIVE AND METHODOLOGY

3.1. OBJECTIVES

To analyze the soil moisture content and predict suitable crops that can be grown in a particular region based on Macronutrients (such as Nitrogen, Phosphorus, and Potassium), temperature, humidity, pH, and rainfall in the agricultural field.

To select suitable machine learning model for the system, we tend to train, test, and validate various machine learning algorithms. Best algorithm can be chosen by comparing the performance of various algorithms.

To predict and analyze moisture content in the soil using Neural Network Pattern Recognition in MATLAB. To develop a website which has input fields to enter values and which will provide corresponding output for the input.

To predict and analyze moisture content in the soil using Neural Network Pattern Recognition in MATLAB. To develop a website which has input fields to enter values and which will provide corresponding output for the input.

3.2. METHODOLOGIES

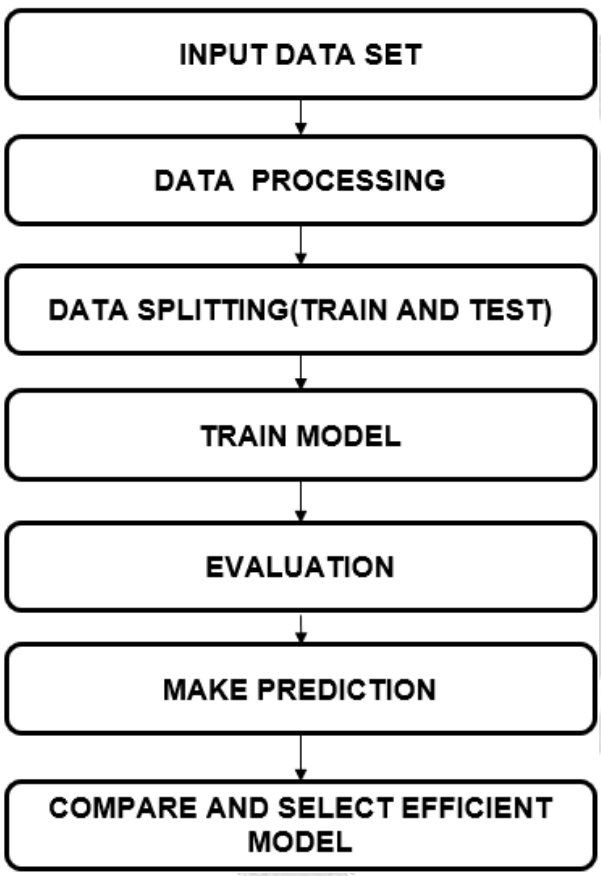


Fig-1: Flowchart for prediction and analysis of soil moisture using machine learning techniques

3.2.1 Input dataset

The input dataset in machine learning is the data used to train machine learning models. Data can be of any type, but its structure is generally easy to process by machine learning algorithms. The quality and quantity of data will directly affect the performance of the model. If data is poorly collected or not organized correctly, then the model will not be able to learn patterns in the data and make accurate predictions.

Our dataset consists of following feature:

1. Nitrogen (N)
2. Phosphorus (P)
3. Potassium (K)
4. Temperature
5. Humidity

6. Ph
7. Rainfall

3.2.2 Data processing

Data preprocessing is the process of preparing raw data for machine learning. It includes processes such as cleaning the data, removing duplicates, and converting the data into a format compatible with machine learning algorithms.

Data preprocessing is an important step in the machine learning pipeline as it can directly impact the performance of the model.

If the data is previously inaccurate, the model will not learn patterns in the data and will not be able to make accurate predictions.

- Removing outliers: Outliers are data points that differ from the rest of the data. Outliers may result from data collection errors or infrequent circumstances. It is important to remove outliers from the dataset before training the model as they may be biased.
- Filling in missing values: Missing values are data points that are not present in the data set. Incomplete results may result from data entry errors or users not wanting to answer certain questions. There are many ways to fill in missing values, such as using the column's mean, median, or mode.
- Converting the data to a compatible format: Machine learning algorithms often require data to be in a specific format. For example, some algorithms expect data to be numerical, while other algorithms require data to be categorical. You may need to convert your data to a compatible format before training your model.

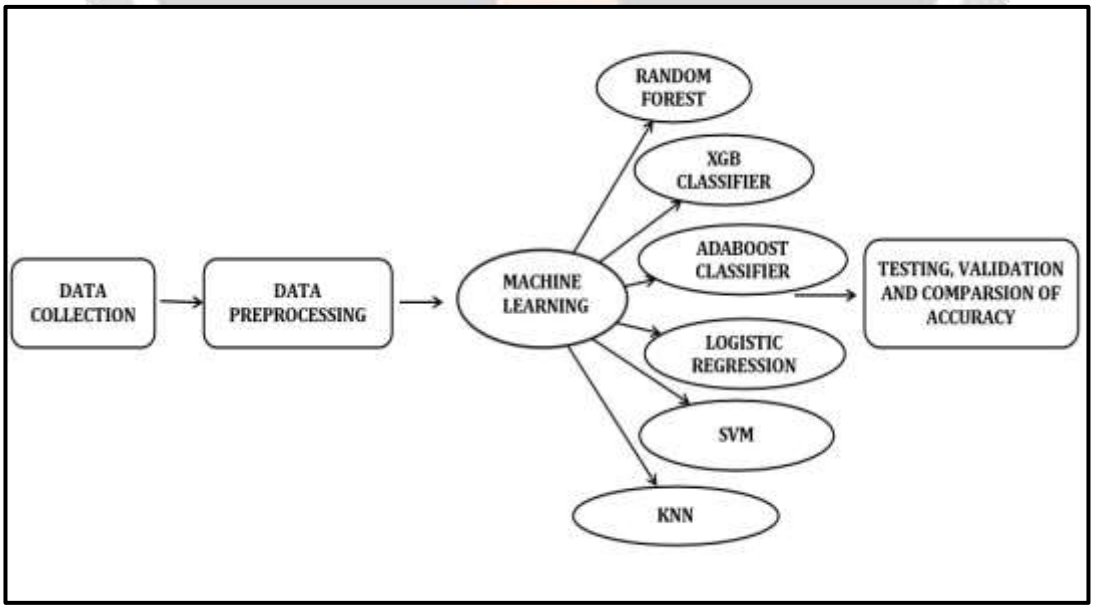
3.2.3 Data splitting

Once the data is clean and ready, you need to split it into three groups: training, validation, and testing. The training process will be used to train your machine learning model. The model will learn patterns in the training data and use those patterns to make predictions on new data. Validation methods will be used to evaluate the performance of the model during training. After each training iteration, you will test your model on the practice set to see how well it performs. This will help you identify problems with your model and adjust as needed.

3.2.4 Train Model:

In machine learning, the training process is the process of teaching a machine learning algorithm to predict target variables based on a set of inputs.

Fig – 2 : Flowchart for prediction and analysis of soil moisture using machine learning techniques



This process involves feeding the algorithm input and matching the values, and then readjusting the parameters of the algorithm until it can predict the target value for the input features.

We are training, testing, and validating the dataset with various machine learning algorithms such as

1. Decision tree
2. AdaBoost classifier

3. XGB classifier
4. Random forest
5. Logistic regression
6. SVM classifier
7. KNN algorithm
8. Neural network pattern recognition

All these algorithms fall into two categories: classification and regression.

Classification:

Classification is a supervised learning task where the goal is to categorize input data into distinct classes or labels based on its features. The algorithm learns from labeled training data, where the input features are mapped to predefined output classes. The objective of a classification algorithm is to accurately predict the class label of new, unseen data points.

3.2.5 Evaluation:

Once the training model is complete, it is important to evaluate the model on retained data to understand how it performs on new data. This will help ensure that the model does not overfit the data. You can use various metrics to evaluate your model, such as accuracy, precision, recall, and F1 score. The best measure to use depends on the specific problem you are trying to solve.

3.2.6 Make Prediction:

Once the model is trained and evaluated, you can start making predictions about new features. To make a prediction, feed new data into the training model and get a model prediction for the target variable. The prediction process in machine learning is a simple process. However, it is worth noting that the quality of the prediction will depend on the quality of the data and the machine learning algorithm used. Use multiple functions.

4. PROPOSED WORK:

4.1. IMPLEMENTATION OF SUPPORT VECTOR MACHINE(SVM)

Support vector machine (SVM) is a supervised machine learning algorithm that can be used for classification and regression. But SVM is mainly used for task classification. SVM works by finding a general plane that divides data points in each area into two groups with as many edges as possible. A hyperplane is a defined boundary that separates points into two groups. Support Vector Machine implementation for the crop_dataset. The crop_dataset is Microsoft Excel .csv (Comma Separated Value) file.

4.2. IMPLEMENTATION OF K NEAREST NEIGHBOR(KNN)

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm that can be used for classification and regression. It is a simple but effective algorithm that can be used to estimate the target difference of new data by finding K similar points in the training data and the determined half of their target values. To train a KNN model, the algorithm needs to be provided with training data that includes input and target variables. Input data can be any data type, such as numeric data, categorical data, or text data. KNN Algorithm implementation for the crop_dataset. The crop_dataset is Microsoft Excel .csv (Comma Separated Value) file.

4.3. IMPLEMENTATION OF LOGISTIC REGRESSION

Logistic regression is a supervised machine learning algorithm that can be used for classification tasks. This is a simple but powerful algorithm that can be used to predict the outcome of a binary event, such as whether a customer will abandon, whether an email is spam, or whether a patient has an illness. Logistic Regression implementation for the crop dataset. The crop dataset is Microsoft Excel .csv (Comma Separated Value) file.

5. RESULTS AND CONCLUSION:

The aim of our project is to design a front-end website. User needs to provide the input in the input panels. After processing, Output will be displayed to the user. We have developed the website using HTML, CSS, and Java script. By using neural network pattern recognition, we have predicted the moisture content in the soil. In addition to that we have also analyzed the macro nutrients present in the soil. Based on the analysis we have predicted the suitable crop that can be cultivated in the region. For prediction we have trained, validated, and tested seven machine learning algorithms. Out of those seven machine learning algorithms we got better accuracies in logistic regression, SVM classifier and KNN classifier.

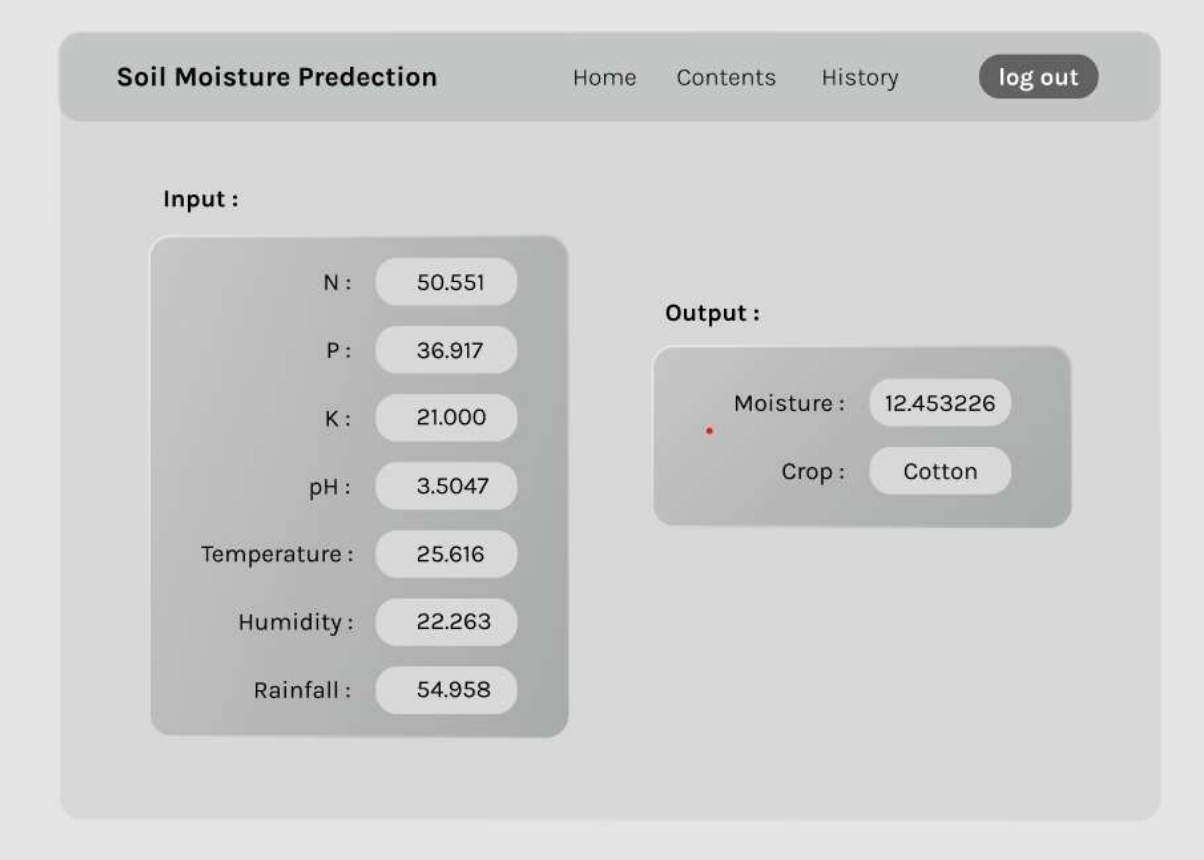


Fig – 3: Website Design

6. REFERENCES:

[1] Anguraj, K., Thiyaneswaran, B., Megashree, G., Shri, J. P., Navya, S., & Jayanthi, J. (2021). Crop recommendation on analyzing soil using machine learning. Turkish Journal of Computer and Mathematics Education, 12(6), 1784-1791.

[2] Senapaty, M. K., Ray, A., & Padhy, N. (2023). IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture. Computers, 12(3), 61.

[3] Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop Prediction Model Using Machine Learning Algorithms. Applied Sciences, 13(16), 9288.

[4] Farwa, U. E., Rehman, A. U., Khan, S. Q., & Khurram, M. (2020). Prediction of Soil Macronutrients Using Machine Learning Algorithm. International Journal of Computer (IJC), 38(1), 1-14.

[5] Ikram, A., Aslam, W., Aziz, R. H. H., Noor, F., Mallah, G. A., Ikram, S., ... & Ullah, I. (2022). Crop Yield Maximization Using an IoT-Based Smart Decision. Journal of Sensors, 2022, 1-15.

[5] Vasu, D., Singh, S. K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V. P., ... & Kalaiselvi, B. (2017). Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management. Soil and Tillage Research, 169, 25-34.

[6] Waiker, V., Pal, M. V., & Gawande, M. A. (2018). Organization learning by analysing Manufacturing losses-A case study.

[7] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.

[8] Ubalanka, V., Jose, A., & Viswanath, D. (2021, May). Machine Learning Strategies for Predicting Crop Diseases. In Journal of Physics: Conference Series (Vol. 1850, No. 1, p. 012119). IOP Publishing.

[9] Mehta, P., Shah, H., Kori, V., Vikani, V., Shukla, S., & Shenoy, M. (2015, March). Survey of unsupervised machine learning algorithms on precision agricultural data. In 2015 international conference on innovations in information, embedded and communication systems (ICIIECS) (pp. 1-8). IEEE.

[10] Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018). A novel kNN algorithm with data-driven k

parameter computation. Pattern Recognition Letters, 109, 44-54.

[11] Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of translational medicine, 4(11).

[12] Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

[13] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Computers and Electronics in Agriculture, 177, 105709.

[14] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." In 2020 third international conference on smart systems and inventive technology (ICSSIT), pp. 926-932. IEEE, 2020.

[15] Ip, R. H., Ang, L. M., Seng, K. P., Broster, J. C., & Pratley, J. E. (2018). Big data and machine learning for crop protection. Computers and Electronics in Agriculture, 151, 376-383.