Adoption of Climate-Smart Agriculture in Farm Production: A Machine Learning Approach

Victoria Tanoh
North Carolina Agricultural and Technical State University

Terence Jude Wood
North Carolina Agricultural and Technical State University

Productivity gains in agriculture are crucial for economic and employment development, yet understanding how they interact is still evolving. Agricultural productivity, particularly crop production, is expected to increase by more than 60% to prevent a global food crisis by 2050. Increased farm production and productivity require inputs and technical services. The farming sector must meet global food demands amid weather changes and unexpected health crises.

Agricultural production must increase by about 70% to cater to food needs in the coming years. This paper utilizes machine learning applications in crop production as an alternative to the current agricultural systems in the United States. Applying machine learning techniques ensures agricultural productivity and is a fruitful step toward mitigating the possibility of a global food crisis.

Keywords: agricultural productivity, machine learning, food crisis, economic development

INTRODUCTION

Food is a basic human need at the center of our dinner tables. However, the sustainability of its production has recently become uncertain and complicated due to the increasing food demand by the rising global population. The current world population is 7.98 billion and is estimated to hit 8 billion by November 2022, as reported by the United Nations. We live in a world currently unsure about food availability and accessibility by 2050, simply because the global population will likely increase to about 9 billion. The farming sector must meet global food demands amid weather changes and unexpected health crises. Agricultural production must increase by about 70% to cater to food needs in the coming years.(Fusco et al., 2020). As old as it sounds, agriculture is one of the industries yet to go through a complete metamorphosis in today's changing world.

Agricultural Productivity

Productivity gains in agriculture are crucial for economic and employment development, yet understanding how they interact is still evolving. Increased farm production and productivity require inputs and technical services. These services are needed to grow crops and process, store, and transport produce, which creates new jobs and growth in the off-farm stages of the agri-food system (AFS). Moreover, raising farm incomes through increased agricultural productivity improves purchasing power, stimulating employment and output growth in non-farm sectors, especially in value-added food chains and non-tradable goods and services (Barrett et al., 2020). However, very few organic inputs such as animal manure, crop residues, and cover crops are used by smallholder farmers partly because such inputs are challenging to utilize in fertility-depleted soils. The agricultural sector plays a significant role in the United States economy, contributing.

Overview of Climate Smart Agriculture

CSA is at the intersection of agricultural science, environmental sustainability, and data science. It involves utilizing innovative techniques, including using the internet, mobile phones, drones, remote sensing equipment, computers, and servers of farm practices. The fundamental strategy of CSA is a careful blend of two crucial United Nations Sustainability Development

Goals- UNSDG, namely, goal 12- responsible consumption and production, and goal 13-Climate Action(Org, n.d.).

The research aims to identify and build a machine-learning model to optimize current farming decisions while ensuring sustainability. The application of machine learning is prevalent in various fields, such as health services, transportation, finance, supply chain, and biotechnology. Researchers and analysts have recently incorporated it into agriculture to increase yield (Benos et al., 2021). Also, machine learning in agriculture is relevant to help small farmers, ranchers, and agricultural analysts strategize to meet global food demand (Liakos et al., 2018). This study will be a foundational model for research and analysts to improve modeling machine learning techniques in agriculture. By replicating the source codes and enhancing the study's limitations, future research and models deployed could be more robust and have better predictive performance. Incorporating data science techniques, machine learning, and the deep neural network recently gained much attention and yielded positive results. The benefit of CSA in crop production, as highlighted by (Liakos et al., 2018), include but are not limited to the following: Crop Production, Crop Mapping, Yield Prediction, Water Management, Soil Management, Weed detection, Pest infestation recognition and Plant species recognition.

Again, this further provides real-time conclusions to farmers using robust predictive machine learning algorithms and analysis to assist farmers in making meaningful decisions. These critical decisions make a huge difference in the consequential actions of farmers on the environment. For instance, a farmer with adequate knowledge of the precise quantity of fertilizer, pesticides, and other chemicals will influence his decision on production inputs. These decisions will contribute immensely to increased yield and the prevention of excess chemicals which otherwise may seep down the soil into the groundwater table. The positive effect of this knowledge could prevent water pollution, which could severely threaten human and animal lives.

Adopting machine learning in agriculture will enhance agricultural productivity. It will be a step toward building resiliency and a robust agricultural sector. Finally, this will instill confidence in the global supply chain regardless of future uncertainty.

Goals

The specific goals of the research are:

- To identify the optimal soil needs of twenty-two cash crops grown in the U.S.
- To build an effective predictive model with high model accuracy that can anticipate the optimal crop given certain soil condititions.

Problem Statement and Justification

Farmers, ranchers, and other key industry players involved in the global supply chain ensure the food on the grocery aisles and our tables. Nevertheless, these key players are yet to benefit entirely from the 21st-century smart, intelligent innovations. Regardless of the rise in global population, farmers have had to bear the weight of food production to feed all. Also, sustainability has become more prevalent due to global warming, unpredictable weather patterns, unexpected droughts, and rising sea levels leading to floods.

Moreover, over-reliance on fertilizer and pesticides has made this once simple task quite daunting for farmers. Hence, the need for efficient agricultural productivity to increase food supply. The overarching problem is climate change and its consequential effect on agriculture. In a spiral motion, climate change affects the temperature, including rising sea levels which causes flooding and extreme droughts in other areas. This change also leads to the adverse effect of salinity intrusion in freshwater bodies, land, and different regions.

Traditional farm practices have been successful but ineffective in meeting global demands.

The effect of unlikely weather patterns has exposed the vulnerabilities in these farm practices. Consequently, there is a need to optimize agricultural production efficiency and enforce environmental safety. At the tail end of this is the greater danger these factors pose to human survival through food insecurity.

The remainder of the paper is organized as follows: Section 2 provides a detailed review of recent studies that have applied machine learning in agriculture and related studies, while section 3 explains the methodology's procedures for employing machine learning. In section4, the results and performance of the Model built are expounded. Finally, the paper's contribution is concluded, and future research is explained.

Framework of Conventional Agricultural Methods

Many frameworks exist regarding traditional farming systems derived from the World Agricultural Systems. Conventional agricultural methods include a range of non-exhaustive farm operations, including; planting, harvesting, pest management, fertilizer application, land tillage, and irrigation, with little attention to environmental sustainability (Devkota et al., 2019). These operations tend to be intensive farm systems with considerable damage to the land. Various literature has attempted to define conventional agricultural methods as any other farm practice that utilizes agrochemicals. However, Chandra et al. (2018) argue that the conventional agricultural system interpretation is unfair. Sumberg and Giller (2022) investigated the definition and context of traditional farm farming systems. The study involved a meta-analysis of various literature focusing on conventional agrarian farming systems. The study concludes that the authors agree that productivity improvements made to traditional practices would enhance the overview of the future of agriculture.

Alternative Farm Methods

Tal (2018) investigated the available sustainable alternatives to traditional farm practices while illuminating the potentially adverse impacts of conventional methods on the environment and human health. The study involved the meta-analysis of 15 studies focusing on environmental sustainability. The authors concluded that adopting organic farming was a suitable substitute for conventional farming, admitting that organic farming may not be optimal. In related studies, Kazimierczak et al. (2019) defined *organic farming* as systems that solely utilize natural procedures in agriculture. Further on, they proposed the following alternative farming methods to mitigate the environmental challenges of conventional farming; organic farming systems, green evolution, and climate-smart agriculture.

Climate Smart Agriculture

Climate Smart Agriculture (CSA) is a framework that encompasses an all-inclusive system in agriculture centered around intentionally increasing agricultural productivity without compromising sustainability. Various researchers have thoroughly assessed the evolution of CSA as an alternative to existing farm systems, including Edward (2020). Chandra (2018) emphasized the need for a well-defined context of CSA without misclassifying it as precision agriculture or intelligent agriculture. The review highlighted the difference between the three farming systems in their approaches. While precision farming purely centers on improving inputs (Chandra et al., 2018), smart farming enhances the entire farming system, from inputs to farm products and beyond CSA (Mazetto et al., 2020). Smart agriculture, however, incorporates external data sources such as analysis of the weather and market trends for optimization. Thus, CSA is a part of Smart farming with three distinct core elements.

The systematic analysis of literature from 2004 to 2016 by Totin et al. (2018) revealed climate vulnerability and food security as the underlying themes for developing CSA models globally. Nevertheless, the three pillars of CSA, namely, productivity, adaptation, and mitigation, must be carefully integrated to achieve global food security. The conclusion of the study further revealed the limitations of CSA as with any other model. The constraints of CSA include the widely subjective nature of the areas adopted, predominantly influenced by available resources and infrastructure in the geographic location and the existing agrarian farm systems. The conclusion of the work further revealed the limitations of CSA as with any other model (Totin et al., 2018).

Machine Learning in Agriculture

The role of agriculture as the backbone of every economy shows the importance of the agriculture industry. Population increases have suddenly and implicitly placed considerable pressure on agricultural productivity to meet the demands of the current population. Modern technology is required to avert laborious work and increase production to address the high need for food. Current research trends indicate a steady surge of the research interest of the scientific research community in agricultural production and the application of machine learning in agriculture. The evolution of ML in scientific research studies began in 1995. However, applying the technique in agriculture gained popularity and highlighted an increasing slope between 2011 and 2019 (Muniasamy, 2020). The resultant popularity stems from the ability of ML to solve several challenges in agricultural operations, including growth deficiencies in fruits such as strawberries, pesticide management, and soil content predictability (Saleem et al., 2022; Chen et al., 2019). Liakos et al. (2018) present a comprehensive review of research on machine learning applications in agricultural production systems. Their study shows ML as a solid predictive analytical technique for agriculture. However, Kamilaris and PrenafetaBoldu (2018), in a comprehensive assessment of 40 studies, argue that deep learning provides a better potential to solve agricultural and food production challenges. Meshram et al. (2021) provide a detailed survey assessing the machine-learning techniques applied in 48 independent studies. These studies combined numerous machine-learning and neural networks in different agricultural production stages. Pathan et al. (2020) summarize 16 studies that applied a combination of available agriculture intelligence such as precision farming, deep learning, image processing, deep learning, and convolution neural network.

Benefit of Machine Learning in Agriculture

Traditional agricultural production methods rely on prescheduled activities without implementing artificial intelligence. Sharma et al. (2021) define Machine learning as a technique under A.I. that utilizes computer systems to study, learn and adapt to patterns and instructions for statistical inference and decision-making. The potential merit of Machine Learning (ML) in agriculture is abundant with evolving techniques (Elavarasan et al., 2018). Elavarasan et al. (2018) distinguished ML as a technique heavily involving vast amounts of data for assessment in a comparative evaluation between ML and statistical modeling utilized under traditional agricultural operations. The application of machine learning is prevalent in various fields, such as health services, transportation, finance, supply chain, and biotechnology. Researchers and analysts have recently incorporated it into agriculture to increase yield (Benos et al., 2021). Also, machine learning in agriculture is relevant to help small farmers, ranchers, and agricultural analysts strategize to meet global food demand. Meshram et al. (2021) summarized the main agricultural-related stages where the adoption of ML suited a pictorial view. The settings are pre-harvesting, harvesting, and post-harvesting. Activities within the preharvesting steps include soil management and monitoring. Also, the harvesting stages consist of crop mapping and disease detection. Finally, post-harvesting is predictive analytics of agricultural yield and productivity.

Soil Management. Yang et al. (2019) perform an intrinsic assessment of four machine-learning approaches by sampling 523 soil samples from various parts of China to deepen the accuracy of ML techniques. A combination of the Nonlinear Extreme Learning Machine (ELM) coupled with a genetic algorithm had the best soil properties prediction accuracy for paddy soils. Suchitra and Pai (2020) applied ELM, an emerging ML technique, to test and classify soil pH and fertility index to enhance environmental sustainability. The Model accurately classified the soil pH and yielded a 90% performance accuracy on the soil fertility indices. The study illuminated the improved ability of ML as a valuable tool in predictive analysis in agriculture.

Disease Detection. Disease detection is one of the phenomenal applications of machine learning in agricultural operations. Although this area of machine learning is still underdeveloped relative to deep understanding, advanced studies combine machine and deep knowledge to enhance disease prediction accuracy. Negi et al. (2021) developed a disease detection model using multiple image datasets to categorize various plant diseases. The Model utilized a deep learning method for disease plant classification. Sindhu et al. (2021) employed ML in disease classification identification in rice plant production.

Harvesting and Post-Harvesting. Harvesting and post-harvesting periods are very relevant stages in agriculture operations. In a review of ML adoption in these areas, Chlingaryan et al. (2018) discuss research developments within the last 15 years on machine learning-based techniques for accurate crop yield prediction and nitrogen status estimation. ML is proven to assist in reducing harvesting and post-harvesting losses through precise image classification. Fatima et al. (2020) applied several machine-learning techniques to four fruit image classifications to automate the sorting process during harvesting. The work showed that ML inbuilt color models have approximately 96% color classification accuracy. For instance, (Garcia et al., 2019) developed an automatic tomato ripeness identification model from 900 tomato images collected from farms and online sources. The tomato images represented various levels of tomato development. Consequentially, the automation of tomato classification was accurate using a support vector machine (SVM) classifier via a machine learning approach and image processing techniques. Thus, throughout literature and continuous research, machine learning has yielded impressive results in the agricultural industry. Furthermore, robust models in machine learning are being developed through constant research and development to cater to evolving issues in the global agricultural supply chain.

METHODS AND DATA

This section provides a detailed description of the theoretical framework of machine learning, the proposed Model, the data sets, and the data process utilized in the preliminary analysis.

Machine Learning Techniques

Machine learning (ML) constitutes utilizing computer systems to study, learn and adapt to patterns and instructions for statistical inference and decision-making -(Sharma et al., 2021). Simply put, it is training a computer to recognize repeating patterns to make classification and predictive analysis in a similar instance. Therefore, the key is to define the problem that will influence data collection critically. The general ML process can be categorized into three stages, as in figure 1 in the appendix. Data Collection and Preprocessing, ML algorithm, Model Selection, and Deployment are the stages.

Stage One: Data Collection and Cleaning

The first stage in the machine learning process is collecting and carefully assessing the type of data collected. Data Preprocessing refers to all the procedures to ensure that the data is clean. They include removing duplicated data and replacing missing numbers with the means. The above process combination is termed exploratory data analysis (EAD). The EAD is to understand the data attributes better and provide quick insight into the data quality. Followed closely is the undertaking of exploratory analysis. Data exploration creates a snapshot of data distribution and visualization, facilitating early error detection. Once the data is cleaned, it is ready for ML processing. The correlation among attributes is estimated to determine multicollinearity.

Stage Two: Data Processing

This stage involves summarizing and visualizing the cleaned dataset. At this stage, the dataset is segregated into two unequal parts: the test data and the train data. For instance, a dataset may be split into 80% training and 20% testing or validation dataset. This approach is known as the train validation test split approach. The idea is to train the Model with enough samples of the training datasets capable of being validated by the test data. This further ensures the accuracy of the Model—finally, the building and establishing of suitable classification models. A key advantage of ML is the versatility of the technique to utilize numerous models to enhance accuracy. About six known ML algorithms are usually exposed to datasets known as ML classifiers. They are Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naïve Bayes (NB), and Support Vector Machines (SVM).

Stage Three: Model Selection and Deployment

The cross-validation method is performed for each possible Model. Thus, this stage consists of selecting the best Model specifically for the dataset with the highest accuracy. The K-fold validation is then used for model selection and deployment. Consequently, the final candidate model is used for classification or a predictive model. Predictions and types are evaluated, and accuracies are assessed.

Proposed Model

The traditional ML model is utilized in our preliminary analysis to yield foundational results and to assess the weakness that may arise from the candidate. The application of ML in agriculture is continually produced. Hence, ML algorithms are applied to time-series agricultural input datasets from 2015 to 2019. Extensive data is collected and segregated in a machine-learning module to train and test data. This process relies on artificial intelligence to finally reach a conclusion or rule after training and testing to improve its prediction accuracy. The elements are phosphorus, nitrogen, and Potassium; the average temperature recorded on crop exposure to sunlight; the soil pH levels; the humidity; the average Rainfall or soil moisture; and finally, the class label (rice, peanut, green beans, and others.).

Data Curation

Due to the limited data availability, the current data set consists of monthly data on U.S. agricultural inputs (Nitrogen, Phosphorus, and Potassium) and necessary production conditions for proper plant growth (pH, rainfall, and temperature). The preliminary information is from the USDA National Agricultural Statistical Service (USDA-NASS) on the top 22 crops grown in the United States from 2015 to 2019. It consists of eight attributes to assist in crop production recommendations for small-scale farmers. Data on fertilizers by nutrient (nitrogen, phosphorus, and Potassium) is accessible from the Food and Agriculture Organization of the United Nations (FAOSTAT), and data on crop production is available from USDA - National Agricultural Statistics Service - Data and Statistics.

Nitrogen is a vital component of amino acids that plants use for food production.

Phosphorus is a necessary element that facilitates plant cell division and tissue formation. Also, Potassium is known as an element that enhances plant functionality. Knowledge of the right temperature, soil acidity or alkalinity level, and rainfall volume is necessary for increasing crop productivity.

RESULTS

While ML is applied in various aspects of agriculture, the research is focused on the preharvesting stage of farming hence adopting CSA practices and ML. Employing artificial intelligence techniques at the pre-planting stage sets the tone for CSA practice. In effect, knowledge of the volume of fertilizer to be applied and the specific soil area to use instead of overall land area application saves time. Most importantly, it goes a long way to ensure that more than necessary fertilizer is not applied to the land. The adverse effect of the latter is the potential of the chemicals to seep down the underground water table regardless of being organic or inorganic. Therefore, the study focused on practical and conscious soil management practices using ML.

Data Preprocessing and Exploratory Analysis

The dataset is investigated for any missing, null, or duplicates in the dataset. The datasets contain 2200 observations with seven attributes and a class label. While seven are the main attributes, and the final column class is the label attribute. Table 1.0 in the appendix shows a detailed description of the crops considered in the study. All analysis and algorithms were performed in Python pandas and Numpy libraries. The dataset showed no missing, null, or duplicates.

The descriptive statistics showed the average volume of fertilizer a farmer uses on 0.5-acre farmland. All elements (N, P K) were in pounds(lb.). Thus, the average nitrogen, phosphorus, and Potassium utilized were approximately 51 lb., 53lb, and 48lb, respectively. Among the three, phosphorus was applied the most, while Potassium was the least used. Also, to eliminate confusion, the element columns are renamed to their traditional names (nitrogen, phosphorus, and Potassium). The dataset is further visualized to provide oversight of the data distribution. Apart from temperature and pH datasets with a gaussian distribution, the remainder of the attributes were non-gaussian. Therefore, they were normalized to translate them to a normal distribution. Figure 2 shows this phenomenon.

Data Processing

Firstly, to answer the first objective stated, the ML algorithm was able to estimate and quantify the top five and least five crops that required any of the above fertilizers and conditions for proper growth. The details of this are in figure 3 below.

The crops that required the most nitrogen were cotton, peanut, tomato, banana, and watermelon. Meanwhile, those that needed the least nitrogen were lentils, pomegranates, oranges, lettuce, and pigeon peas. Also, crops that needed significant phosphorus were apples, grapes, bananas, lentils, and chickpeas, and those needing the least were oranges, cucumbers, watermelons, tomatoes, and pomegranates. Similarly, grapes, apples, chickpeas, watermelon, and tomatoes required the most Potassium, while oranges, peppers, lentils, cotton, and maize required the least Potassium. For the soil conditions, we start with the pH levels; chickpeas, peppers, oranges, lentils, and cotton would thrive in areas with high pH levels regardless, while kidney beans, lettuce, pigeon peas, apple, and cucumber would suffer in a low place. Also, papaya, lettuce, peppers, tomato, and green beans are exposed to need the most sunlight.

In contrast, chickpeas, kidney beans, pomegranate, maize, and apples would not survive extremely high temperatures. Hence, they have the least amount of sunlight. Finally, rice, cucumber, wheat, peanut, and pigeon peas needed the most rainfall or moisture volume. Tomato, lentils, green beans, watermelon, and soybeans needed the least rain.

Model Selection and Deployment

All six ML algorithms were applied to the training dataset for objective two. The train and test datasets were split into 80% and 20%, respectively. Carefully not to overfit the models, the accuracies, recall, f1-score, and support for all models were recorded for selection. Details of the models results are that for the KNN classifier, SVC, Logistic Regression, Random Forest, and the Light Gradient Boost Model (LGBM), the accuracies were 97% (0.97), 98% (0.98), 97% (0.97), 99%(0.99) and 99%(0.99) respectively. Although the random forest and the LGBM both had the same accuracies, the selected Model was the LGBM because it utilizes the least memory while being extremely efficient at optimizing the learning process.

Thus, the selected LGBM model was further trained on various samples of the training dataset. Finally, the Model was applied to the test dataset and accurately predicted the top 5 crops that may be grown using the optimized level of all seven attributes.

DISCUSSION AND FUTURE RESEARCH

ML application undoubtedly has the potential to significantly reduce the challenge of overfertilization of soil and low crop yield. Attempt to solve these issues have resulted in traditional farming methods that have adversely affected the ecosystem. The research has demonstrated the ability of ML to predict the agricultural inputs necessary for optimal productivity accurately. In the preliminary studies, we sought to identify the grey area in agriculture that ML could improve. Numerous aspects of agriculture could benefit from ML.

Future works would consist of a comprehensive farm specific dataset collection from farms in North Carolina. As a collaborative effort between the Green Agritech Foundation, Greensboro, NC, USA, data collected would be analyzed, and a final predictive model would be suggested and employed by farmers on a pilot project geared towards enhancing food security in North Carolina.

ACKNOWLEDGEMENT

This work was proudly sponsored by the Green Agritech Foundation, Greensboro, NC.

REFERENCES

Barrett, C.B., Reardon, T., Swinnen, J., & Zilberman, D. (2020). Agri-food value chain revolutions in low-and middle-income countries. *Journal of Economic Literature*, *58*, 1–67.

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors (Basel, Switzerland)*, *21*(11). https://doi.org/10.3390/S21113758

Chandra, A., McNamara, K.E., & Dargusch, P. (2018). Climate-smart agriculture: Perspectives and framings. *Climate Policy*, *18*(4), 526–541.

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.

Chen, Y., Lee, W.S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., & He, Y. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing*, *11*(13), 1584.

Devkota, K.P., Pasuquin, E., Elmido-Mabilangan, A., Dikitanan, R., Singleton, G.R., Stuart, A.M., . . . Beebout, S.E. (2019). Economic and environmental indicators of sustainable rice cultivation: A comparison across intensive irrigated rice cropping systems in six Asian countries. *Ecological Indicators*, *105*, 199–214.

Edwards, C.A. (2020). *Sustainable agricultural systems*. CRC Press.

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, *155*, 257–282.

Fatima, S., & Seshashayee, M. (2020, December). Hybrid color feature image categorization using machine learning. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 687–692). IEEE.

Garcia, M.B., Ambat, S., & Adao, R.T. (2019, November). Tomayto, tomahto: A machine learning approach for tomato ripening stage identification using pixel-based color image classification. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (pp. 1–6). IEEE.

Kamilaris, A., & Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, *147*, 70–90.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

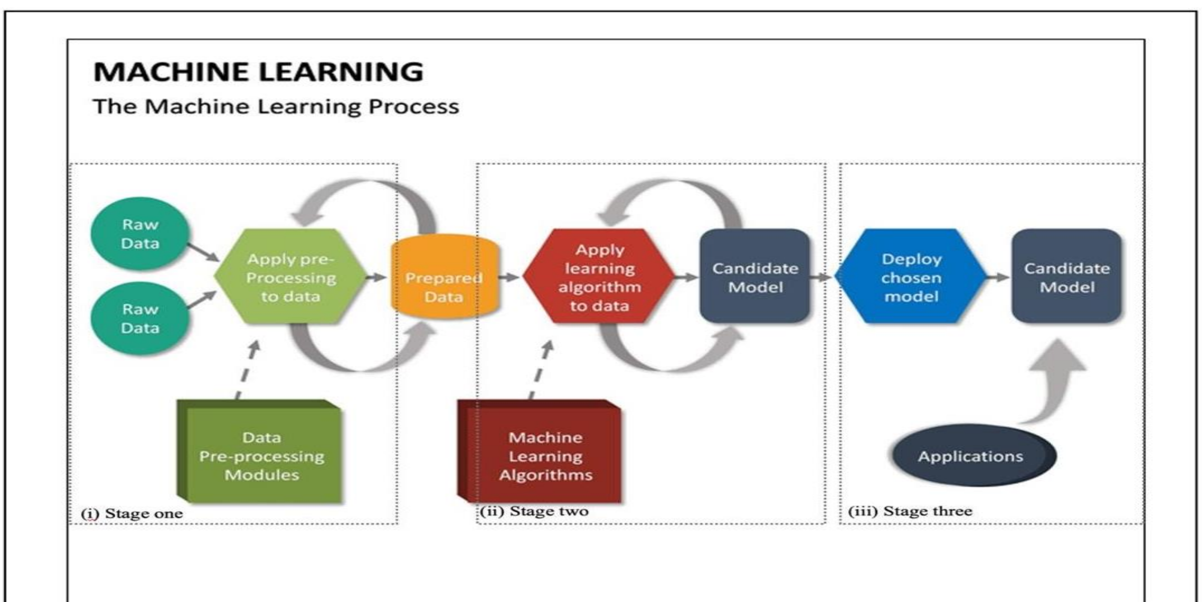
Mazzetto, F., Gallo, R., & Sacco, P. (2020). Reflections and methodological proposals to treat the concept of "information precision" in intelligent agriculture practices. *Sensors*, *20*(10), 2847.

Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S.D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010.

Negi, A., Kumar, K., & Chauhan, P. (2021). Deep neural network-based multi-class image classification for plant diseases. *Agricultural Informatics: Automation Using the IoT and Machine Learning*, pp. 117–129.
Pathan, M., Patel, N., Yagnik, H., & Shah, M. (2020). Artificial cognition for applications in intelligent agriculture: A comprehensive review. *Artificial Intelligence in Agriculture*, *4*, 81–95.
Rehman, T.U., Mahmud, M.S., Chang, Y.K., Jin, J., & Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture*, *156*, 585–605.
Saleem, M.H., Potgieter, J., & Arif, K.M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precision Agriculture*, *22*(6), 2053–2091.
Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, *9*, 4843–4873.
Sindhu, P., Indirani, G., & Dinadayalan, P. (2021). Fuzzy C-Means (FCM) Clustering with Probabilistic Neural Network (PNN) model for detection and classification of rice plant diseases in the internet of things-cloud centric precision agriculture. *Journal of Computational and Theoretical Nanoscience*, *18*(4), 1194–1200.
Suchithra, M.S., & Pai, M.L. (2020). Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Information processing in Agriculture*, *7*(1), 72–82.
Sumberg, J., & Giller, K.E. (2022). What is 'conventional'agriculture?. *Global Food Security*, *32*, 100617.
Totin, E., Segnon, A.C., Schut, M., Affognon, H., Zougmoré, R.B., Rosenstock, T., & Thornton, P.K. (2018). Institutional perspectives of climate-smart agriculture: A systematic literature review. *Sustainability*, *10*(6), 1990.
Yang, M., Xu, D., Chen, S., Li, H., & Shi, Z. (2019). Evaluation of machine learning approaches to predict soil organic matter and pH using Vis-NIR spectra. *Sensors*, *19*(2), 263.

APPENDIX

FIGURE 1
THE MACHINE LEARNING MODEL



Machine Learning Process

FIGURE 2
CORRELATION MATRIX OF ATTRIBUTES

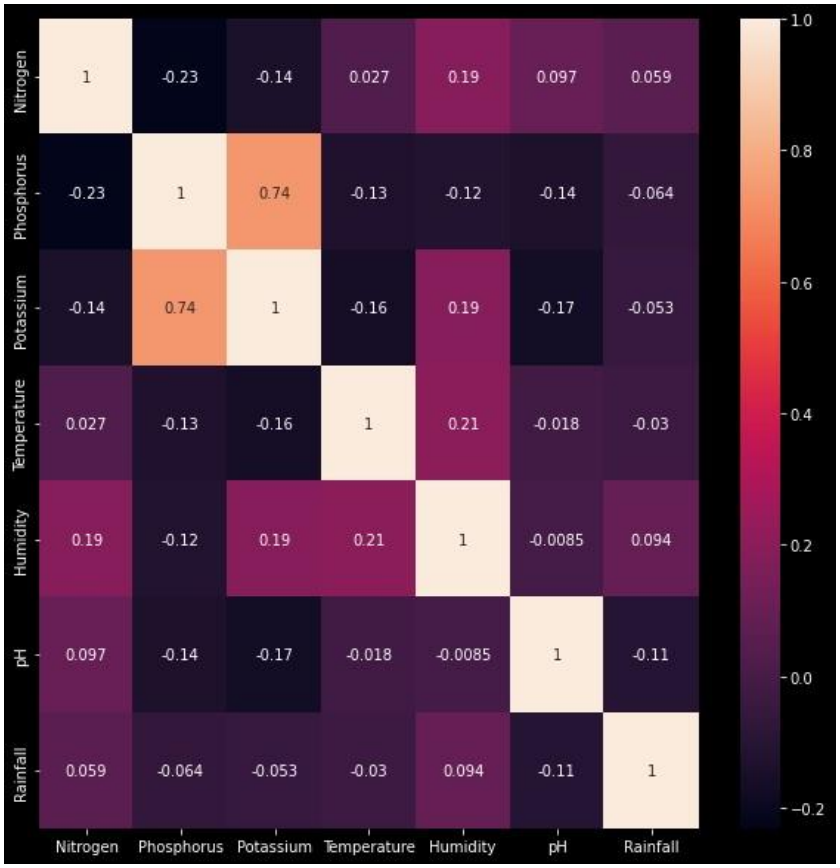


FIGURE 3(A) AND (B)

```
--------------------------------
Top 5 Least Nitrogen requiring crops:
--------------------------------
lentil --> 18.77
pomegranate --> 18.87
orange --> 19.58
lettuce --> 20.07
pigeonpeas --> 20.73
--------------------------------
Top 5 Least Phosphorus requiring crops:
--------------------------------
orange --> 16.55
cucumber --> 16.93
watermelon --> 17.0
tomato --> 17.72
pomegranate --> 18.75
--------------------------------
Top 5 Least Potassium requiring crops:
--------------------------------
orange --> 10.01
peppers --> 19.24
lentil --> 19.41
cotton --> 19.56
maize --> 19.79
--------------------------------
Top 5 Least Temperature requiring crops:
--------------------------------
chickpea --> 18.8728467519
kidneybeans --> 20.1150846851
pomegranate --> 21.837841721999997
maize --> 22.3892039102
apple --> 22.6309424132
--------------------------------
Top 5 Least Humidity requiring crops:
--------------------------------
chickpea --> 16.8604394237
kidneybeans --> 21.6053567295
pigeonpeas --> 48.0616330847
lettuce --> 50.1565726953
soybeans --> 53.16041802790001
--------------------------------
Top 5 Least pH requiring crops:
--------------------------------
kidneybeans --> 5.749410585870001
lettuce --> 5.766372799660001
pigeonpeas --> 5.794174879790001
apple --> 5.929662931809999
cucumber --> 5.97656212619
--------------------------------
Top 5 Least Rainfall requiring crops:
--------------------------------
tomato --> 24.689952066
lentil --> 45.680454204
green bean --> 48.403600902899996
watermelon --> 50.7862189449
soybeans --> 51.198487045700006
--------------------------------
```

```
--------------------------------
Top 5 Most Nitrogen requiring crops:
--------------------------------
cotton --> 117.77
peanut --> 101.2
tomato --> 100.32
banana --> 100.23
watermelon --> 99.42
--------------------------------
Top 5 Most Phosphorus requiring crops:
--------------------------------
apple --> 134.22
grapes --> 132.53
banana --> 82.01
lentil --> 68.36
chickpea --> 67.79
--------------------------------
Top 5 Most Potassium requiring crops:
--------------------------------
grapes --> 200.11
apple --> 199.89
chickpea --> 79.92
watermelon --> 50.22
tomato --> 50.08
--------------------------------
Top 5 Most Temperature requiring crops:
--------------------------------
papaya --> 33.7238587388
lettuce --> 31.2087701513
peppers --> 29.9733396789
tomato --> 28.663065756
green bean --> 28.5257747353
--------------------------------
Top 5 Most Humidity requiring crops:
--------------------------------
cucumber --> 94.84427180610001
papaya --> 92.4033876826
tomato --> 92.34280196089999
apple --> 92.3333828756
orange --> 92.17020876340001
--------------------------------
Top 5 Most pH requiring crops:
--------------------------------
chickpea --> 7.33695662374
peppers --> 7.13395162948
orange --> 7.01695745276
lentil --> 6.927931571609999
cotton --> 6.91267549578
--------------------------------
Top 5 Most Rainfall requiring crops:
--------------------------------
rice --> 236.18111359399998
cucumber --> 175.686645804
wheat --> 174.792797536
peanut --> 158.066294882
pigeonpeas --> 149.4575638135
--------------------------------
```

FIGURE 4-7
ML ALGORITHMS, RESPECTIVELY

```
import warnings
warnings.filterwarnings('ignore')
param_grid={'C':[0.001,0.01,0.1,1,10,100], 'max_iter':[50,75,100,200,300,400,500,700]}
log=RandomizedSearchCV(LogisticRegression(solver='lbfgs'),param_grid,cv=5)
log.fit(X_train,y_train)
y_pred_log=log.predict(X_test)
confusion_log=confusion_matrix(y_test,log.predict(X_test))
plt.figure(figsize=(8,8))
sns.heatmap(confusion_log,annot=True)
plt.xlabel("Predicted")
plt.ylabel("Actual")
print(classification_report(y_test,y_pred_log))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       1.00      1.00      1.00        30
           3       1.00      1.00      1.00        30
           4       1.00      1.00      1.00        30
           5       1.00      1.00      1.00        30
           6       0.97      1.00      0.98        30
           7       1.00      0.97      0.98        30
           8       0.88      1.00      0.94        30
           9       1.00      1.00      1.00        30
          10       1.00      1.00      1.00        30
          11       1.00      1.00      1.00        30
          12       1.00      0.93      0.97        30
          13       1.00      1.00      1.00        30
          14       0.97      0.97      0.97        30
          15       0.94      0.97      0.95        30
          16       1.00      1.00      1.00        30
          17       0.88      0.77      0.82        30
          18       1.00      0.87      0.93        30
          19       0.97      1.00      0.98        30
          20       1.00      1.00      1.00        30
          21       0.79      0.90      0.84        30

    accuracy                           0.97       660
   macro avg       0.97      0.97      0.97       660
weighted avg       0.97      0.97      0.97       660
```

Logistic regression

```
[ ] param_grid = {
    'n_estimators': [50, 75,100, 150, 200,300],
    }
    rcv=RandomizedSearchCV(RandomForestClassifier(random_state=42),param_grid,cv=5)
    rcv.fit(X_train,y_train)
    y_pred_rcv=rcv.predict(X_test)
    confusion_rcv=confusion_matrix(y_test,rcv.predict(X_test))
    plt.figure(figsize=(8,8))
    sns.heatmap(confusion_rcv,annot=True)
    plt.xlabel("Predicted")
    plt.ylabel("Actual")
    print(classification_report(y_test,y_pred_rcv))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       1.00      1.00      1.00        30
           3       0.97      1.00      0.98        30
           4       1.00      1.00      1.00        30
           5       1.00      1.00      1.00        30
           6       1.00      1.00      1.00        30
           7       1.00      1.00      1.00        30
           8       1.00      1.00      1.00        30
           9       1.00      1.00      1.00        30
          10       1.00      0.97      0.98        30
          11       1.00      1.00      1.00        30
          12       1.00      1.00      1.00        30
          13       1.00      1.00      1.00        30
          14       1.00      1.00      1.00        30
          15       1.00      1.00      1.00        30
          16       1.00      1.00      1.00        30
          17       1.00      0.90      0.95        30
          18       1.00      1.00      1.00        30
          19       1.00      1.00      1.00        30
          20       1.00      1.00      1.00        30
          21       0.91      1.00      0.95        30

    accuracy                           0.99       660
   macro avg       0.99      0.99      0.99       660
weighted avg       0.99      0.99      0.99       660
```

```
[ ] param_grid={'C':[0.001,0.01,0.1,1,10,100], 'gamma':[0.001,0.01,0.1,1,10,100]}
    rcv=RandomizedSearchCV(SVC(),param_grid,cv=5)
    rcv.fit(X_train,y_train)
    y_pred_svc=rcv.predict(X_test)
    confusion_svc=confusion_matrix(y_test,rcv.predict(X_test))
    plt.figure(figsize=(8,8))
    sns.heatmap(confusion_svc,annot=True)
    plt.xlabel("Predicted")
    plt.ylabel("Actual")
    print(classification_report(y_test,y_pred_svc))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       1.00      1.00      1.00        30
           3       1.00      1.00      1.00        30
           4       1.00      1.00      1.00        30
           5       1.00      1.00      1.00        30
           6       1.00      1.00      1.00        30
           7       1.00      1.00      1.00        30
           8       0.91      1.00      0.95        30
           9       1.00      1.00      1.00        30
          10       1.00      1.00      1.00        30
          11       1.00      1.00      1.00        30
          12       1.00      1.00      1.00        30
          13       1.00      0.97      0.98        30
          14       1.00      0.97      0.98        30
          15       1.00      1.00      1.00        30
          16       1.00      1.00      1.00        30
          17       0.92      0.80      0.86        30
          18       1.00      0.93      0.97        30
          19       1.00      1.00      1.00        30
          20       1.00      1.00      1.00        30
          21       0.80      0.93      0.86        30

    accuracy                           0.98       660
   macro avg       0.98      0.98      0.98       660
weighted avg       0.98      0.98      0.98       660
```

```python
import lightgbm as lgb
model = lgb.LGBMClassifier()
model.fit(X_train, y_train)
y_pred=model.predict(X_test)
confusion=confusion_matrix(y_test,y_pred)
plt.figure(figsize=(8,8))
sns.heatmap(confusion,annot=True)
plt.xlabel("Predicted")
plt.ylabel("Actual")
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       1.00      1.00      1.00        30
           3       0.97      1.00      0.98        30
           4       1.00      1.00      1.00        30
           5       1.00      1.00      1.00        30
           6       1.00      1.00      1.00        30
           7       1.00      1.00      1.00        30
           8       1.00      1.00      1.00        30
           9       0.97      1.00      0.98        30
          10       1.00      0.97      0.98        30
          11       1.00      1.00      1.00        30
          12       1.00      1.00      1.00        30
          13       1.00      1.00      1.00        30
          14       0.97      1.00      0.98        30
          15       1.00      0.93      0.97        30
          16       1.00      1.00      1.00        30
          17       0.93      0.90      0.92        30
          18       1.00      1.00      1.00        30
          19       1.00      1.00      1.00        30
          20       1.00      1.00      1.00        30
          21       0.90      0.93      0.92        30

    accuracy                           0.99       660
   macro avg       0.99      0.99      0.99       660
weighted avg       0.99      0.99      0.99       660
```

```python
knn=KNeighborsClassifier(n_neighbors=1)
knn.fit(X_train,y_train)
from sklearn.metrics import confusion_matrix
confusion_knn=confusion_matrix(y_test,knn.predict(X_test))
plt.figure(figsize=(8,8))
sns.heatmap(confusion_knn,annot=True)
plt.xlabel("Predicted")
plt.ylabel("Actual")
from sklearn.metrics import classification_report
print(classification_report(y_test,knn.predict(X_test)))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       1.00      1.00      1.00        30
           3       0.97      1.00      0.98        30
           4       0.97      1.00      0.98        30
           5       1.00      1.00      1.00        30
           6       1.00      1.00      1.00        30
           7       1.00      1.00      1.00        30
           8       0.87      0.90      0.89        30
           9       0.97      1.00      0.98        30
          10       1.00      0.97      0.98        30
          11       1.00      0.97      0.98        30
          12       1.00      1.00      1.00        30
          13       1.00      0.97      0.98        30
          14       0.88      0.97      0.92        30
          15       1.00      1.00      1.00        30
          16       1.00      1.00      1.00        30
          17       0.89      0.80      0.84        30
          18       1.00      0.83      0.91        30
          19       1.00      1.00      1.00        30
          20       1.00      1.00      1.00        30
          21       0.79      0.90      0.84        30

    accuracy                           0.97       660
   macro avg       0.97      0.97      0.97       660
weighted avg       0.97      0.97      0.97       660
```

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
X_train=scaler.fit_transform(X_train)
X_train=pd.DataFrame(X_train,columns=X.columns)
X_train.head()
```

	Nitrogen	Phosphorus	Potassium	Temperature	Humidity	pH	Rainfall
0	0.688592	-0.157057	-0.594893	-1.249816	-0.084923	-0.883905	-0.365413
1	1.794585	0.872296	-0.062038	-0.076920	0.349076	-1.092737	0.038137
2	-1.361541	-0.974484	-0.318598	1.996698	-0.775947	-0.038057	-0.202136
3	-0.579254	-0.580908	-0.437010	1.602002	-0.769901	-0.240009	-0.254929
4	-0.822033	-1.398335	-0.062038	-1.326020	0.796273	-0.493530	0.166280

```
X_test=scaler.transform(X_test)
X_test=pd.DataFrame(X_test,columns=X.columns)
X_test.head()
```

	Nitrogen	Phosphorus	Potassium	Temperature	Humidity	pH	Rainfall
0	-1.307590	-0.520358	-0.338333	1.019727	-0.969211	-0.941132	-0.209803
1	1.309027	-1.035034	-0.298863	0.361583	-0.719448	-0.240283	0.496265
2	0.041181	0.085144	0.727376	-1.088866	-2.565505	0.462324	-0.266007
3	-1.118762	0.085144	-0.614628	0.087952	-0.213548	-0.255546	-1.032692
4	-1.091787	0.781470	-0.555422	-0.137740	-0.208043	-0.103042	-1.182928

```
from sklearn.svm import SVC
from sklearn.model_selection import RandomizedSearchCV
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
knn_scores=[]
for k in range(1,20):
    knn=KNeighborsClassifier(n_neighbors=k)
    scores=cross_val_score(knn,X_train,y_train,cv=5)
    knn_scores.append(scores.mean())

x_ticks = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]
x_labels = x_ticks

plt.plot([k for k in range(1,20)],knn_scores)
plt.xticks(ticks=x_ticks, labels=x_labels)
plt.grid()
```

Link to thegoogle collaboratory document for source codes:
https://colab.research.google.com/drive/1679dKDzzEwRCJpQuOx1b0kokgPPWxtYg#scrollTo=Dpazs78uzVz9