Agrisense: Agricultural Support System Based On Machine Learning For Farmers

Arjun Haridas
Department of Electronics Engineering
Sardar Patel Institute of Technology
Mumbai, India
arjun.haridas@spit.ac.in

Pranjal Agrawal
Department of Electronics Engineering
Sardar Patel Institute of Technology
Mumbai, India
pranjal.agrawal@spit.ac.in

Namooh Shah
Department of Electronics Engineering
Sardar Patel Institute of Technology
Mumbai, India
namooh.shah@spit.ac.in

***Abstract*— The Indian agricultural landscape is characterized by diverse agro-climatic conditions, fragmented landholdings, and a growing demand for sustainable farming practices. To address these challenges, we propose an intelligent agricultural support system that leverages the power of data analytics to provide farmers with tailored crop recommendations, plant disease identification, and fertilizer suggestions. Our system is built upon a comprehensive dataset collected through surveys from farmers across India, capturing their unique agricultural experiences and challenges.**

***Keywords*— Crop recommendation, Machine learning,XGBoost,Plant disease identification,Fertilizer**

I. Introduction

An Agricultural Support System (ASS) plays a pivotal role in bolstering the efficiency, productivity, and sustainability of the agricultural sector. As the global population continues to burgeon, the demand for food surges in tandem, underscoring the critical importance of modernizing and fortifying agricultural practices. An effective Agricultural Support System is designed to provide comprehensive assistance to farmers, agribusinesses, and stakeholders involved in the agricultural value chain.

At its core, an Agricultural Support System encompasses a multifaceted approach, integrating technology, information dissemination, financial support, and skill development. It serves as a linchpin for farmers by equipping them with the tools and knowledge necessary to optimize crop yields, mitigate risks, and adopt sustainable farming practices. This system acts as a nexus, connecting farmers with the latest advancements in agricultural science, weather forecasting, market trends, and government policies, thereby empowering them to make informed decisions.

Key components of an Agricultural Support System include access to cutting-edge technology such as precision farming techniques, drones, and smart sensors. These technologies aid in precision agriculture, enabling farmers to optimize resource use, minimize environmental impact, and enhance overall productivity. Additionally, financial support mechanisms, including subsidies, loans, and insurance programs, are integral to mitigating economic uncertainties for farmers.

Moreover, the Agricultural Support System serves as a conduit for knowledge transfer and skill development. Training programs and workshops facilitate the dissemination of best practices, fostering a culture of innovation and adaptability among farmers. By promoting sustainable farming methods, the system contributes to the long-term resilience of agriculture in the face of evolving environmental challenges.

An effective Agricultural Support System is the cornerstone of a thriving and resilient agricultural sector. By leveraging technology, knowledge, and financial resources, it not only uplifts the livelihoods of farmers but also contributes to global food security and sustainable development. As we navigate the complexities of the 21st century, a robust Agricultural Support System stands as a beacon, guiding the agricultural community towards a more prosperous and sustainable future.

Our project aims to incorporate three applications based on the common problems faced by the farmers in our country which are Crop recommendation, Fertilizer recommendation, and Plant disease prediction.

II. Methodology

The proposed method predicts the yield of a crop based on both climatic and geographical data as well as the composition of the soil of that particular region. The sequential model which is a Deep Learning Model is created and trained with the above data to forecast the yield. A crop recommendation system is also designed to suggest a suitable crop for the soil based on the soil nutrients using different machine -learning techniques to maximize the crop yield

A. Dataset Collection

Diverse datasets, including soil characteristics, climate patterns, and crop disease images, were collated from reliable sources. Farmer preferences and challenges were captured through surveys distributed on social media platforms, establishing a robust foundation for system development. The climate data and geographical data were gathered from different government websites. There are several Internet summaries accessible. The sample dataset used for yield prediction is shown in Table 1. The crop recommendation dataset consists of nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall values which are collected from kaggle.com. For the plant disease detection, a huge resource of existing plant disease dataset was used with several thousand plant disease images present.

TABLE 1. Crop recommendation sample data

N	P	K	Temperature(C)	Humidity %	pH	Rainfall(mm)	Label
90	42	43	21	82.00274423	6.502985292	202.9355362	Rice
85	58	41	22	80.31964408	7.038096361	226.6555374	Rice
60	55	40	23	82.3207629	7.840207144	263.9642476	Rice
74	35	44	25	80.15836264	6.980400905	242.8640342	Rice

a.) *Climate Data*

The Indian Meteorological Department (IMD) provides daily, monthly, and annual reviews of climate factors in India, including minimum temperature, maximum temperature, apparent temperature, dew point, precipitation, latitude, longitude, pressure, wind speed, visibility, cloud cover, etc. Kaggle also contains different datasets regarding climate and soil.

b.) *Geographic Data*

NASA and other space organizations provide data about the earth's surface.The majority of vegetation indices are data condensed, which enables them to precisely and economically characterize natural vegetation cover and development conditions.

For disease prediction, we employed ResNet, a deep convolutional neural network (CNN) architecture. ResNet was trained on a large dataset of crop images labeled with corresponding plant diseases. The trained model can effectively identify various plant diseases based on uploaded images of affected crops. The dataset consists of about 87K rgb images of healthy and diseased crop leaves which is categorized into 38 different classes. The total dataset is divided into 80/20 ratio of training and validation set preserving the directory structure. A new directory containing 33 test images is created later for prediction purpose

Figure 1. Plant diseases

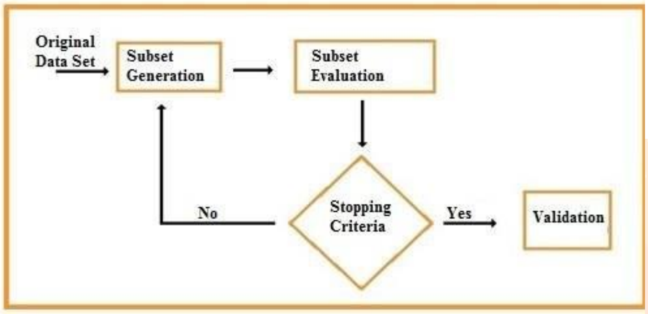
B. Data Cleaning

When integrating several data sources, there are several opportunities for data to be mislabeled or duplicated. Inaccurate or partial data might cause operational efficiency to be disrupted. The practice of eliminating inaccurate, corrupted, improperly organized, redundant, or insufficient information from a dataset is known as data cleaning. Prior to usage, the data are preprocessed. It describes every change made to the raw data prior to feeding it into the deep learning model. Additionally, preprocessing is required to speed up training. The obtained data typically contain a significant amount of noise from multiple sources. It is possible to turn messy data into clean data by using data preparation. A data pre-processing technique called data smoothing eliminates noise from a data set.

a.) *Feature Selection*

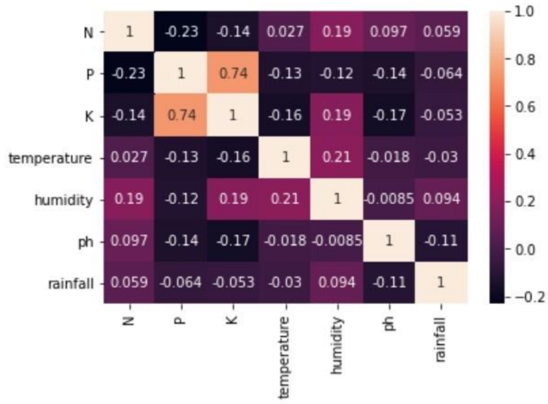
As shown in Figure 2, feature selection is performed to choose the important features before training the model. Additionally, it helps identify traits that aren't relevant, which lessens overfitting and might enhance performance. In addition, when a model has fewer variables and is less complex overall. It is the process of selecting, either automatically or manually, the subset of the most useful and accurate attributes to be utilized in the creation of the model.

Figure 2. Feature selection flowchart



Supervised and unsupervised feature selection models are the two different categories of models. The three types of supervised models are filter, wrapper, and embedding approaches. In this instance, filtering methods are used. Features are eliminated in this manner according to how they relate to the output or even the result. We utilize correlation to determine whether the features should be eliminated if there is a positive or negative relationship between them and the output labels. A subset of features from latitude, longitude, maximum temperature, minimum temperature, apparent temperature, humidity, pressure, and NDVI is chosen for crop yield from a list of attributes in a dataset. As seen in Figure 3, crop recommendations take into account rainfall, pH levels, N, P, and K.

Figure 3. Features for crop recommendation

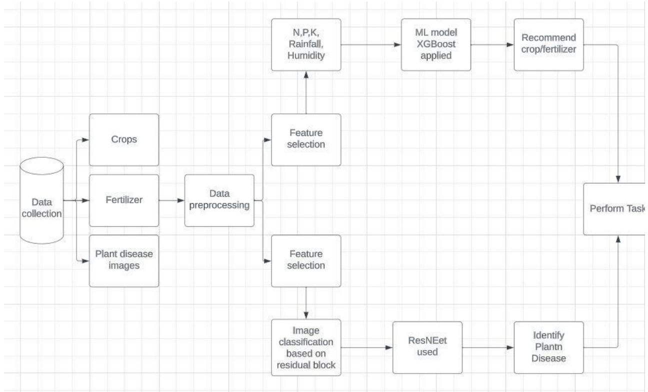


C. *Train the model and predict crop yield*

For a straightforward layer series where each layer has a single input and output tensor, a sequential method is appropriate. Deep learning neural network models are frequently trained using the stochastic gradient descent-optimization method. An optimizer is a software or method that modifies a neural network's properties, such as the weights and learning rate. Ultimately, this reduces overall loss and improves accuracy. Using stochastic gradient descent, we employ a random selection of data batches in place of the original data for each cycle.

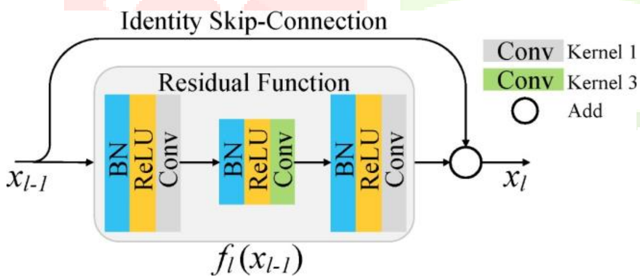
Figure 4 represents the block diagram of the model that is proposed. It includes all the procedures needed to build and train the model, accurately predict crop production, and suggest a crop based on the parameters provided, such as N,P, K, humidity, and rainfall using various algorithms

Figure 4. Block diagram



For disease prediction, ResNet is used which have been one of the major breakthroughs in computer vision since they were introduced in 2015.n ResNets, unlike in traditional neural networks, networks, where each layer feeds into the one below it and into layers that are two to three hops away directly. We employ a network with residual blocks. and directly into the layers about 2–3 hops away, to avoid over-fitting (a situation when validation loss stops decreasing at a point and then keeps increasing while training loss still decreases). This also helps in preventing vanishing gradient problems and allows us to train deep neural networks. Here is a simple residual block:

Figure 5. Residual Block



D. *Machine learning techniques*

a.) Decision Tree

The entire sample space is iteratively divided into smaller sub-sample spaces by the decision tree technique, each of which can be expressed by a simple model. The root node of the tree contains the full sample space.The process of splitting a sample space into several subspaces entails splitting the root node into child nodes, each of which is then continuously split into leaf nodes. That is a graphic representation of how to find every potential solution to the problem based on specific characteristics. To create a tree, we use the CART technique, which stands for the Classification and Regression Tree algorithm.

b.) Naive Bayes

Based on the Bayes theorem One of the simplest and most effective ways to create quick predictive models that can produce quick predictions is the naive Bayes technique. Given that it's a classification algorithm, predictions are inferred based on the likelihood of an entity. Following data pre-processing, we would fit the Naive Bayes classifier to training data. Typically, a Gaussian NB classifier is used to adjust it to the training set. Various classifiers yielded an event B from an event A. It is very easy to carry out. For training, it requires less data than before. could be employed based on our requirements as well.

c.)Support Vector Machine

SVM is used to address problems with classification and regression. The goal of the SVM method was to identify the optimal line and decision boundary for classifying n-dimensional space so that additional data points could be easily added to the correct group later on. A hyperplane is a boundary that stands for the best course of action. By increasing the margin to the maximum between support vectors, this algorithm finds the best differentiating classifier between two classes by utilizing kernel functions such as polynomial, linear, and others.

d.)Logistic Regression

Using a range of characteristics, observations can be categorized using logistic regression, which can also be used to quickly identify the variables that will be most helpful for classification. It is a type of regression that makes use of the predictive analysis theory.

Instead of fitting a regression line, we fit a logistic function with a "S" structure in logistic regression, which forecasts two maximum values, 0 and 1. An equation in mathematics used to convert expected values into probabilities is called the sigmoid function.

e.)Random Forest

One kind of machine learning technique that can be used to address regression and classification issues is called Random Forest. It is based on ensemble methods, which combine multiple classifiers to solve complex problems and enhance the performance of the model. The random forest classifier combines several methods of the same kind.The output of the forest is frequently assessed using the value of the average prediction made by several trees. The random tree is the same as C4.5 or CART, with the exception that it selects a random subset of features. The efficiency increases and the risk of overfitting is reduced with increasing tree density in the forest.It takes less time to train than other approaches.

f.) XGBoost

An ensemble model technique called XGBoost is used. A method for combining multiple learners' prediction performance in an organized manner is called ensemble learning. Ensemble modeling is a helpful technique for raising the performance of a model. The foundational learners, or models, of the ensemble may come from different learning algorithms or from the same ones. Gradient boosted decision trees are created in XGBoost. Each of the independent factors is given a weight, which is subsequently fed into the decision tree to predict results.

E. Finding the best model to recommend crop and fertilizer

The parameters of the trained model are assessed, and useful predictive values are given. The model's performance can be assessed using metrics such as root mean square error, mean squared error, mean absolute percentage error, and mean absolute error. The mean squared error (MSE) measure is used to evaluate the accuracy and loss of a trained model. The recommendation system achieves an accuracy of 99.31%. When compared to alternative algorithms, XGBoost performs better in emphasizing functional space and lowers model costs. The accuracy of the different algorithms used for crop recommendation is shown in Table 2. The yield plot of fertilizer obtained is shownin Figure 6.

Figure 6. Yield plot of fertilizer

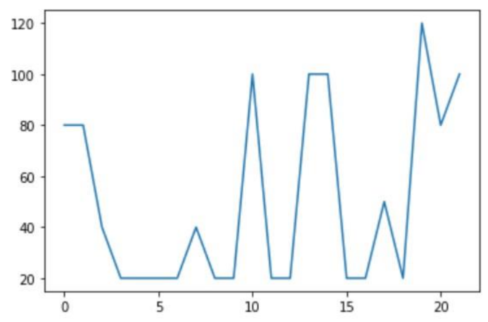


Figure 7 shows that the accuracy achieved with XgBoost is higher than that of other machine learning techniques, such as SVM, Decision Tree, Linear Regression, Naïve Bayes, RF, and XGBoost.

Figure 7. Accuracy of various models

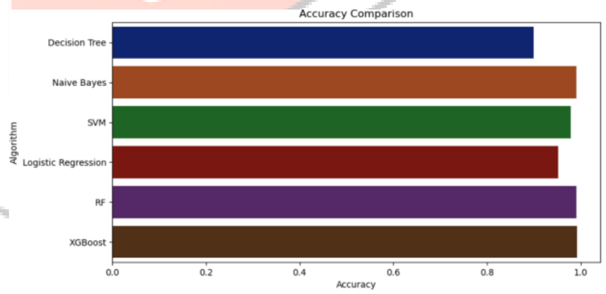


TABLE 2. Accuracy of various models for crop andfertilizer recommendation

Algorithm	Accuracy
DecisionTree	90%
NaïveBayes	99.09%
SupportVectorMachine	97.95%
LinearRegression	95.22%
RandomForest	99.09%
xgboost	99.31%

F. *Plant disease classification technique selected*

When working with image datasets, it is best to use GPU rather than CPU because GPUs are better suited for training deep learning models than CPUs because they can handle multiple computations at once. Their large number of cores makes it possible to compute multiple parallel processes more effectively. Furthermore, deep learning computations must manage enormous volumes of data, which is best served by a GPU's memory bandwidth. We define two helper functions (get_default_device & to_device) and a helper class DeviceDataLoader to move our model and data to the GPU as needed in order to use a GPU seamlessly, if one is available.

Figure 8 shows the image per class of each plant disease

Image classification is then used on the residual block and the class functions are:

training_step: To determine the model's "wrongness" following the validation or training step.We are using this function in addition to an accuracy metric, which is probably not going to be differentiable (that is, it would be impossible to determine the gradient, which is required for the model to get better during training).

Taking a brief glance at the PyTorch documentation reveals the cross_entropy cost function.

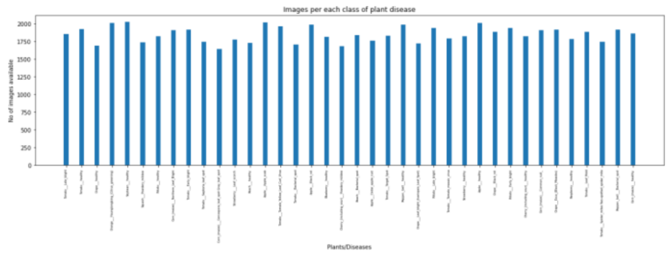
validation_step: Just because an accuracy metric isn't applicable during model training doesn't mean it shouldn't be used! In this instance, accuracy would be determined by setting a threshold and counting the number of times the difference between the predicted label and the actual label by the model is less than the threshold.

validation_epoch_end: After every epoch, we wish to track the train losses and validation losses/accuracies. However, each time we do this, we must ensure that the gradient is not being tracked.

epoch_end: Since we are using a learning rate scheduler, which modifies the learning rate following each training batch, we also wish to print validation losses/accuracies, train losses, and learning rate after each epoch.

We also define an accuracy function which calculates the overall accuracy of the model on an entire batch of outputs, so that we can use it as a metric in fit_one_cycle

Figure 8. Image per class of each plant disease



III. EXPERIMENTAL RESULTS

After being trained, the algorithm can predict crop production with accuracy and suggest the best crop. The model developed has a 99.3 percent accuracy rate for crops in all seasons throughout India. A Deep Neural Network sequential model allows us to calculate the agricultural output based on multiple factors. Latitude, longitude, apparent temperature, pressure, humidity, temperature maximum, temperature minimum, and NDVI are the input parameters that our model uses. The user enters these parameters in a web page interface that shows the fertilizer to be used (as shown in Figure 11). A crop recommender makes recommendations for a good crop using the XGBoost method. Even with a small agricultural area, farmers can still reap the benefits of the suggested approach.

Figure 9. Services offered to the farmer

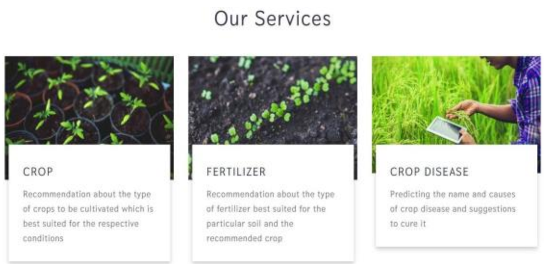


Figure 10. Input values to be entered by farmer

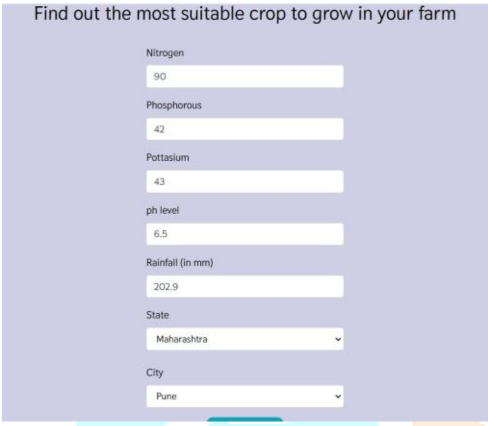


Figure 11. Crop recommendation by the system

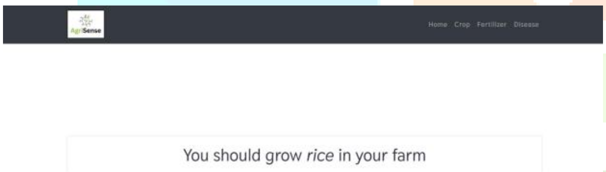
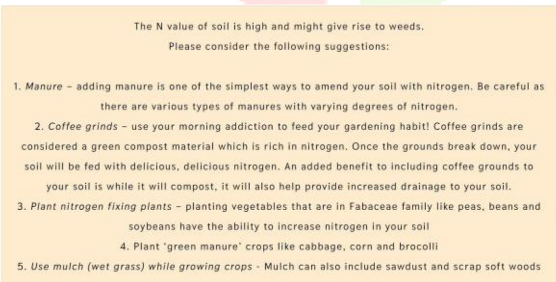


Figure 12. Fertilizer recommendation by the system



The N value of soil is high and might give rise to weeds.
Please consider the following suggestions:

1. *Manure* – adding manure is one of the simplest ways to amend your soil with nitrogen. Be careful as there are various types of manures with varying degrees of nitrogen.
2. *Coffee grinds* – use your morning addiction to feed your gardening habit! Coffee grinds are considered a green compost material which is rich in nitrogen. Once the grounds break down, your soil will be fed with delicious, delicious nitrogen. An added benefit to including coffee grounds to your soil is while it will compost, it will also help provide increased drainage to your soil.
3. *Plant nitrogen fixing plants* – planting vegetables that are in Fabaceae family like peas, beans and soybeans have the ability to increase nitrogen in your soil
4. Plant 'green manure' crops like cabbage, corn and brocolli
5. *Use mulch (wet grass) while growing crops* - Mulch can also include sawdust and scrap soft woods

Figure 13. Plant disease photo to be uploaded by farmer

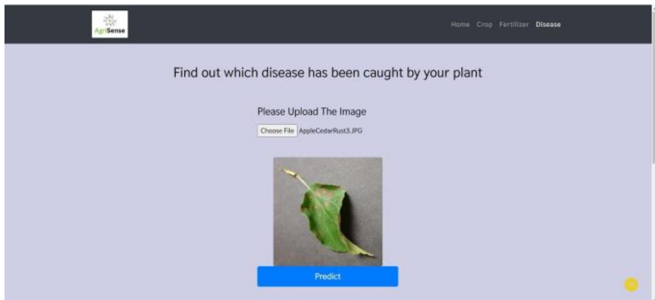
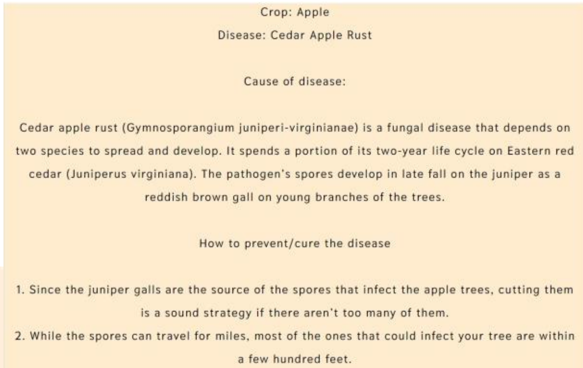


Figure 14. Plant disease identification and prevention methods suggested by the system



Crop: Apple
Disease: Cedar Apple Rust

Cause of disease:

Cedar apple rust (Gymnosporangium juniperi-virginianae) is a fungal disease that depends on two species to spread and develop. It spends a portion of its two-year life cycle on Eastern red cedar (Juniperus virginiana). The pathogen's spores develop in late fall on the juniper as a reddish brown gall on young branches of the trees.

How to prevent/cure the disease

1. Since the juniper galls are the source of the spores that infect the apple trees, cutting them is a sound strategy if there aren't too many of them.
2. While the spores can travel for miles, most of the ones that could infect your tree are within a few hundred feet.

IV. FUTURE SCOPE

Future iterations will expand the system's capabilities, incorporating real-time weather updates and market trends. Continued collaboration with farming communities will drive ongoing improvements, ensuring the system evolves to meet the dynamic needs of agriculture. Additionally, efforts will be directed towards developing a mobile application for increased accessibility in remote areas. The ongoing refinement of ML models will be a key focus to enhance the system's predictive accuracy and broaden its applicability across diverse agricultural landscapes.

V. CONCLUSION

Our intelligent agricultural support system, AgriSense built upon comprehensive datasets, advanced machine learning algorithms, and a user-friendly web interface, holds immense promise for empowering Indian farmers with data-driven insights and decision-making tools. By enhancing crop productivity, minimizing losses, and promoting sustainable practices, our system has the potential to transform Indian agriculture, leading to increased food production, improved farmer livelihoods, and a more resilient agricultural sector. We envision a future where Indian farmers are equipped with the knowledge and tools to make informed decisions, optimize their agricultural practices, and contribute to a thriving and sustainable agricultural sector.

References

[1]"Digital Revolution in the Agriculture Based on Data Science | IEEE Conference Publication | IEEE Xplore," *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/document/9831580 (accessed Nov. 28, 2023).

[1] Mourtzinis S, Esker PD, Specht JE, Conley SP. Advancing agricultural research using machine learning algorithms. Sci Rep. 2021 Sep 9;11(1):17879. doi: 10.1038/s41598-021-97380-7. PMID: 34504206; PMCID: PMC8429560.

[2] "Big Data Analytics based Smart Agriculture | IEEE Conference Publication| IEEE Xplore," *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/document/9489196 (accessed Nov. 28, 2023).

[3] Kumar, Ashish & Kaur, Jagdeep. (2023). Soil Classification Using Machine Learning, Deep Learning, and Computer Vision: A Review. 10.1007/978-981-19-9876-8_25.

[4] Talaviya, Tanha & Shah, Dhara & Patel, Nivedita & Yagnik, Hiteshri & Shah, Manan. (2020). Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artificial Intelligence in Agriculture. 4. 10.1016/j.aiia.2020.04.002.