International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I10P112

Volume 72 Issue 10, 119-129, October 2024

Original Article

A Smart Agricultural Framework for Soil Image Classification Using Modified DenseNet and Crop Recommendation System Using Random Forest

Sreelata Alugoju[1], P. Praveen[2]

[1]*SR University, Warangal, Telangana, India.*
[2]*School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India.*

[1]*Corresponding Author : sreelataalugoju@gmail.com*

Received: 27 December 2023 Revised: 15 May 2024 Accepted: 8 October 2024 Published: 25 October 2024

Abstract *- The way farming and agricultural practices are conducted has changed as a result of the use of the Internet of Things (IoT) and Artificial Intelligence (AI) technology. The development of smart agricultural systems, where automation and autonomous operations are possible, is made possible by IoT and AI technology. The proposed model's goal is to classify soil images using deep learning methods in order to determine the type of soil. With the help of this proposed modified DenseNet model, an accurate assessment of soil characteristics, including fertility, moisture content, and nutrient levels, will be made possible. Modified DenseNet delivers enhanced feature propagation, effective parameter use, resistance against overfitting, and precise findings when used to classify soil image data. The categorized soil data will be used to create an automated crop recommendation system utilizing the random forest algorithm, together with meteorological data and other pertinent criteria. Multiple decision trees are used in the ensemble learning technique known as Random Forest. It uses the combined wisdom of these trees to provide reliable predictions. The Random Forest method's averaging or voting process reduces the impact of individual trees' biases and faults, producing more reliable crop selection suggestions. Based on specific soil characteristics, this technology will provide farmers with customized recommendations for acceptable crop selections. As a result, it will help farmers improve their land management techniques, allowing them to attain maximum production and sustainable results*.

Keywords *- Soil image classification, Crop recommendation, Artificial Intelligence, Internet of Things, DenseNet, Random Forest, Ensemble learning.*

1. Introduction

Agriculture is one of several fields that have been significantly transformed by technological advancements in areas such as Artificial Intelligence (AI) and the Internet of Things (IoT) [1]. The use of AI and IoT in agricultural settings has created new opportunities for the creation of more efficient systems for classifying soils and recommending crops [2]. These innovations in technology have the potential to significantly enhance agricultural operations, maximize the use of resource availability, and increase crop yields. In this article, Artificial Intelligence (AI) and the Internet of Things (IoT) will be examined in the context of agriculture, with a focus on soil classification and crop recommendation. One of the main areas where AI and IoT have significantly impacted the agricultural sector is the categorization of soil [3]. In the past, the classification of soil was done manually via observation and study, which was time-consuming and prone to human error. Farmers may now utilize advanced sensing technology and machine learning algorithms to automate the process of soil categorization [4]. IoT devices like soil moisture sensors, temperature sensors, and nutrient sensors may be scattered over the land in order to get real-time information on the condition of the soil in agricultural areas [5]. These sensors continuously monitor the different properties of the soil and transmit that data to a centralized system. In order to accurately identify the soil type, the data is then examined by AI algorithms that include factors including the soil's moisture content, nutrient levels, pH value, and texture [6]. This computerized soil categorization enables farmers to choose crops, irrigation strategies, and nutrient management practices wisely, which eventually leads to greater crop growth and higher agricultural output. Additionally, AI-powered systems for classifying soil may provide a wealth of information on the fertility and health of the soil [7]. The ability of these systems to identify patterns and trends helps farmers comprehend the long-term effects of soil conditions on crop productivity. This is accomplished via the analysis of historical data and the correlation of that data with crop performance. Farmers are given the ability to enhance soil

health and maintain optimum growing conditions as a result of this information, which enables them to execute targeted interventions such as crop rotation and soil amendments. The categorization of soil is only one of the ways that artificial intelligence and internet of things technologies have changed crop recommendation systems. These systems use algorithms for machine learning in order to assess massive quantities of data, such as the historical performance of crops, weather patterns, soil properties, and consumer demand. AI algorithms are able to deliver customized suggestions to farmers about crop selections that are suited for certain regions and seasons by taking into consideration the parameters mentioned above [8]. Devices connected to the IoT are crucial in the collection of real-time data for use in crop recommendation systems [9]. The data on climatic conditions, such as temperature, rainfall, humidity, and solar radiation, is continually gathered via the use of technology such as weather stations, satellite images, and remote sensing. When paired with data about the soil that IoT sensors have gathered, this information allows AI algorithms to offer suggestions that are both accurate and timely regarding the best crop choices. Farmers may use these guidelines as a basis for making choices motivated by data, which will allow them to maximize the possible yield and minimize the risks associated with poor meteorological conditions. The use of AI and the Internet of Things in agriculture makes precision agricultural approaches easier to implement [10]. Farmers are able to monitor crop health, identify pests and illnesses at an early stage, and perform targeted treatments thanks to the combination of data from Internet of Things devices, such as drones and robots equipped with cameras and sensors, with artificial intelligence algorithms. This strategy decreases the amount of agrochemicals used, lessens the amount of waste produced, and encourages the adoption of sustainable agricultural techniques.

2. Literature Survey

Tien-Heng Hsieh et al. [11] developed a variety of Convolutional Neural Network (CNN) variations were, to classify Hyperspectral Images (HSIs) of agricultural regions, including 1D-CNN with pixel-wise spectral data, 1D-CNN with specified bands, 1D-CNN with spectral-spatial features, and 2D-CNN with principal components. HSI data from agriculture in the Salinas Valley and mixed vegetation agriculture in Indian Pines were used to evaluate the performance of these CNN algorithms. Zhenrong Du et al. [12] employed Deep semantic segmentation networks to extract CA from high-resolution RS pictures and then classify the results. Using images from WorldView-2 (WV-2) with just the Red-Green-Blue (RGB) bands, it was possible to verify that the proposed semantic classification framework is effective for both the information extraction job and the CA mapping activity. Specifically, the authors developed a platform for sampling, training, and testing, in addition to categorizing, in order to extract and map CA on the basis of DeepLabv3+ by making use of TensorFlow, which is a framework for deep learning. Vittorio Mazzia et al. [13] utilized both Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) to create a new and best deep learning model for pixel-based Land Cover and Crop Classification (LC&CC). They used this model on multi-temporal Sentinel-2 images of central north Italy, a region known for its diverse farming system, which is mostly made up of foods grown for sale. The suggested method gets rid of the need for human feature engineering and crop phenological stage modelling by automating the extraction of features by learning the time association between many shots. In their research, the study looked at 15 groups, including important farming crops. Ce Zhang et al. [14] suggested scale sequence because it replaces the conventional scale selection paradigm by incorporating a series of scales into the iterative process of fitting the joint distribution implicit in the Joint Deep Learning (JDL) approach, the Joint Deep Learning (SS-JDL) strategy for joint LU and LC classification is simple and affordable. The scales enable the sequential transfer of information from low-level characteristics to high-level representations and from simple LC states to complex LU characterizations.

They are separately created and used to establish the CNN input patch sizes. The effectiveness of the special SS-JDL technology was assessed using aerial digital imagery of three intricate and diverse landscapes in Southern England (Bournemouth and Southampton) and Northern England (Manchester). Several LC and LU strategies, including the cutting-edge Joint Deep Learning (JDL) technique, were provided as comparison points. Rahim Azadnia et al. [15] utilized a convolutional neural network (CNN) to classify soil texture photos captured at 20, 40, and 60 cm. This system may be deployed everywhere a smartphone can be used. The proposed CNN model is composed of two parts, each with several layers. The first block (feature extraction) consists of the layers Conv, Max-pooling, drop out, and batch normalization.

The second block (classifier) is composed of the flatten, SVM classifier, and fully connected layers. Pallavi Srivastava et al. [16] explored various computer-based soil categorization techniques in two streams. The first involves techniques based on image processing and computer vision that distinguish between soil using different attributes such as texture, colour, and particle size. These methods use common image processing algorithms and methods. The second category consists of techniques for classifying soil using deep learning and machine learning, such as CNN, which provide state-of-the-art results. Deep learning applications streamline the whole process, which essentially lowers the dependency on spatial-form designs and pre-processing techniques. In this paper, the researchers also provide a number of databases that are in line with the study's objectives. Databases are made using many tools,

including digital cameras, digital camcorders, and smartphone cameras, under a variety of lighting and environmental circumstances. Additionally, a short discussion of assessment metrics is included to set out some graded measurements for distinction. Amit Bhola et al. [17] analyzed how well different Machine Learning (ML) systems do at predicting the right crop depending on soil and meteorological conditions. Random Forest, Decision Tree, Support Vector Machine, Naive Bayes, XGBoost, along K-Nearest Neighbour are six supervised machine learning algorithms that have been implemented and analyzed.

Liheng Zhong et al. [18] provided a categorization system for remotely sensed time series based on deep learning. The study was carried out in Yolo County, California, which has a very diverse irrigation-based agriculture system dominated by cash crops. For the challenging task of classifying summer crops using Landsat Enhanced Vegetation Index (EVI) time series, two different deep learning models were developed: one is based on one-dimensional convolutional (Conv1D) layers, while the other is based on Long Short-Term Memory (LSTM).

For comparison, three well-known classifiers—the Random Forest, the Support Vector Machine, and a gradient-boosting algorithm called XGBoost—were also put to the test. Sequential data is often represented using LSTM. Zeel Doshi et al. [19] presented AgroConsultant, an intelligent system that would aid Indian farmers in selecting the crop to produce by taking into account the sowing season, the location of their farms, the qualities of the soil, and climatic elements like temperature and rainfall. Andreas Kamilaris et al. [20] conducted a study of 40 research projects that use deep learning techniques to solve different problems in agriculture and food supply. The authors look at the specific farming problems that are being studied, the models and frameworks that are being used, the sources, types, and pre-processing of the data that is being used, as well as the general success based on the measures that are being used at each work. They also look at how classification or regression performance is different between deep learning and other famous methods that already exist.

Krupa Patel et al. [21] investigated the many categories of recommendation systems and the domains in which they are used. The authors next investigate alternative evaluation criteria for recommendation systems, followed by unresolved problems and research challenges. They look more closely at the research that has already been done in the field. Our algorithm for suggesting crops based on numerous characteristics has been submitted for the Agriculture sector as part of our contribution via this study. P. Parameswari et al. [22] employed decision trees, SVM, and RNN algorithms to create a hybrid model. This study looks at the information and helps farmers predict the crop, increasing their profitability.

3. Proposed Framework

The proposed approach uses deep learning to categorize soil pictures for the evaluation of farmland. This will make it possible to analyze soil characteristics, including fertility, moisture content, and nutrient levels, with accuracy. An automated crop recommendation system will be created using the identified soil data, weather information, and other characteristics to identify the best crops for a given soil condition. Farmers will be assisted in managing their land more effectively in order to maximize production and sustainability.

3.1. Soil Image Classification

When classifying soil images, machine learning techniques like Convolutional Neural Networks (CNNs) are often used. This comprises the automatic classification and identification of different soil types according to their visual appearance. This technique may be used for a number of things, including assessing the fertility of farmland, recommending crops that are appropriate for the area, and keeping track of how environmental conditions affect the soil's quality. The categorization of soil pictures has many advantages, such as being non-destructive, rapid, and able to differentiate between challenging soil types. It offers efficient soil condition data collection. However, there are challenges to be solved, including the necessity for high-quality photographs for accurate classification, sensitivity to changes in lighting and the environment, and the need for enormous amounts of labelled data in order to build effective classification models.

3.1.1. Proposed DenseNet Model

DenseNet is a densely connected Convolutional Neural Network (CNN) in which each layer is connected to all of the layers that came before it. Because of this, DenseNet is able to train more effectively and avoid the issue of vanishing gradients, which may arise in deep CNNs. The layers used in the proposed DenseNet architecture are:

Input layer

The data are initially processed by a neural network at this layer, which is the first layer of the network. It is a representation of the unprocessed characteristics or data that is sent into the network. The dimensionality of the data that is being entered is reflected in the number of nodes that are present in the input layer.

Convolutional Layer

Convolutional Neural Networks (also known as CNNs) rely heavily on this particular component. Convolutional operations are performed on the input data using a collection of learnable filters or kernels as the transformative component. These filters glide through the input data, carrying out element-wise multiplication and summing, which assists in the identification of patterns and characteristics. Convolutional layers are designed to capture

spatial hierarchies and are commonly used in image recognition tasks.

Batch Normalization

This is a method that is used to enhance the training of neural networks as well as their overall performance. The outputs of the layer below are normalized by this operation, which involves removing the batch mean and dividing the result by the batch standard deviation. This helps to solve the issue of the internal covariate shift, stabilize the training process, and speed up convergence. Batch normalization also functions as a regularizer, which eliminates the need for using additional regularisation methods such as dropout.

ReLU (Rectified Linear Unit)

In neural networks, this is an example of an activation function that is often employed. By outputting the input value if it is positive and 0 otherwise, it creates a non-linear relationship between the two variables. ReLU is computationally efficient and contributes to the solution of the vanishing gradient issue. As a result, neural networks are able to learn complicated associations, which leads to an improvement in model performance.

Max Pooling

Downsampling is a method that is used in CNNs in order to bring the spatial dimensions of the feature maps down to a more manageable level. It then outputs the greatest value inside each zone and separates the input into regions that do not overlap with one another. The most important characteristics may be extracted with the aid of max pooling, which also helps to minimize the network's spatial complexity and add some translation invariance.

Average Pooling

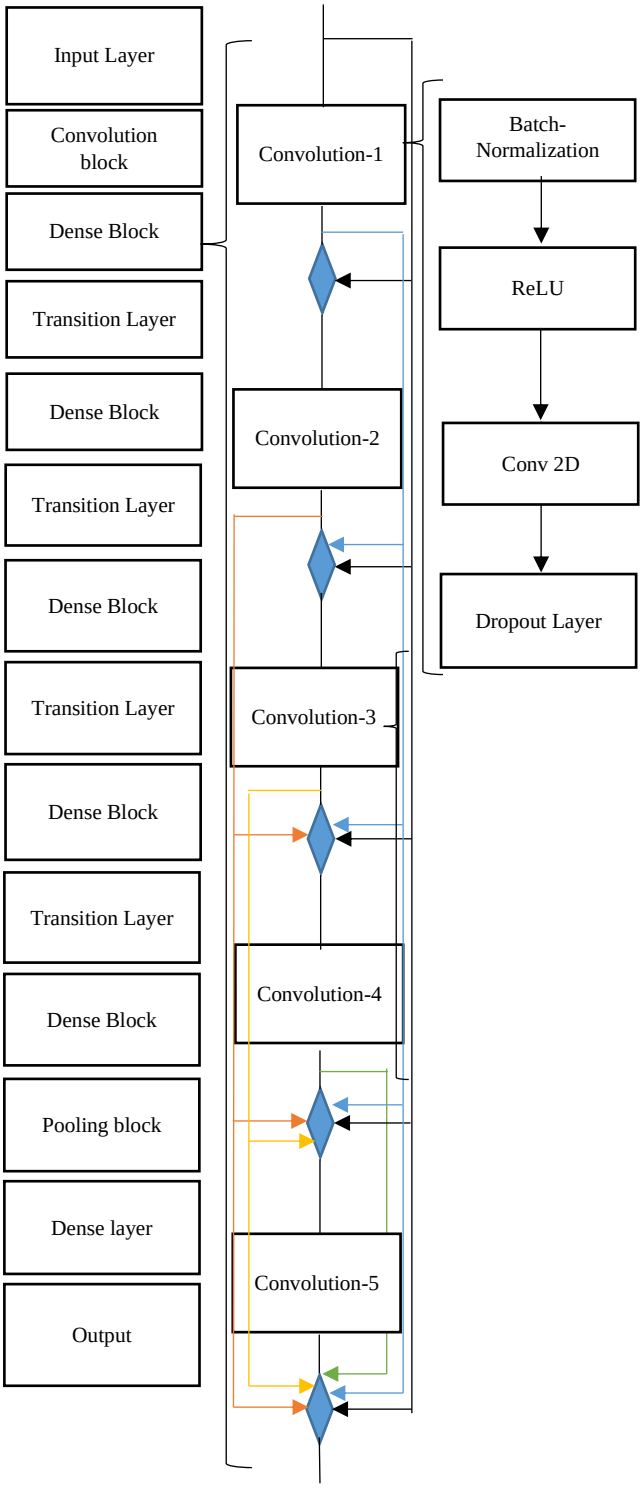
Similar to max pooling, average pooling is another downsampling technique that partitions the input into non-overlapping regions. However, instead of selecting the maximum value, average pooling computes the average value within each region. Average pooling can be useful when spatial localization is less critical, but preserving global information is desired.

Dense Layer

This is a common kind of layer that may be found in neural networks. A dense layer is one in which every node in the layer below is linked to every other node in that layer. The application of weights and biases to the input data, followed by running the resulting data through an activation function, is how dense layers facilitate the learning of complicated patterns.

Dropout

This is a regularisation approach that is used in neural networks to avoid overfitting from occurring. In the process of training, the dropout function will, at random, set a portion of the nodes in a layer to zero, therefore "dropping out" those nodes. This forces the network to learn redundant representations and improves generalization. Dropout helps prevent complex co-adaptations and promotes robustness in the model.



(a) Proposed DensNet architecture

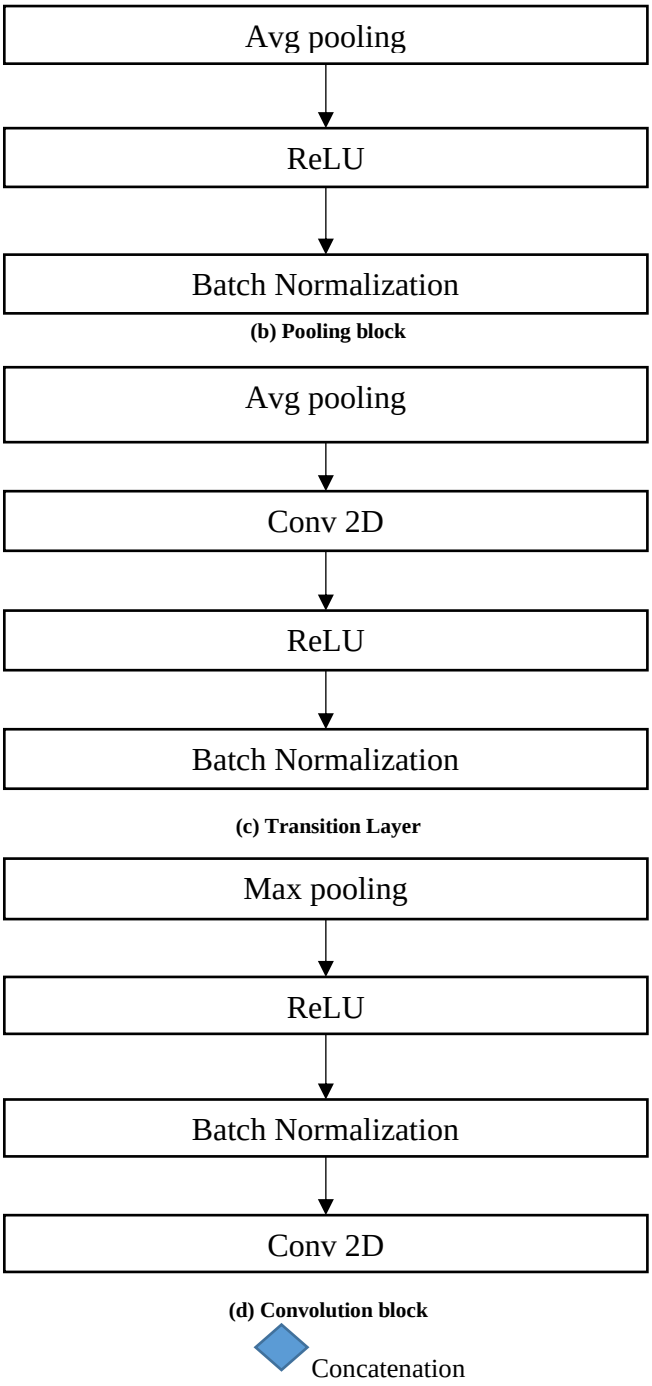


Fig. 1 Proposed DensNet architecture

Traditional DenseNet design defines a "dense block" as a collection of layers that are densely linked to one another. Each succeeding layer in the dense block is provided with inputs from all of the layers that came before it in the block. This dense connection architecture improves feature reuse, encourages gradient flow, and contributes to the alleviation of the issue of disappearing gradients. The dense block that was indicated was adjusted such that it was replaced with a new dense block that included batch normalization, ReLU activation, convolution, and dropout. You said that this was the updated version. The model achieves regularisation, non-linearity, feature extraction, and normalization as a result of the combination of batch normalization, ReLU activation, convolution, and dropout inside the modified dense block. This design is useful for a wide range of deep learning problems because it supports increased training stability, gradient flow, and generalization performance. A transition layer is often used in the traditional DenseNet architecture to downsample the feature maps and lower the spatial dimensions before sending them to the next dense block.

A convolutional layer and a pooling layer, most often a 2D average pooling layer, make up the conventional transition layer. The modified transition layer is made up of average pooling, a 2D convolutional layer, ReLU activation, and batch normalization. The goal of this modification is to keep the information integrity intact while simultaneously lowering the spatial dimensions. This setting guarantees that critical characteristics are kept and efficiently learnt throughout the downsampling process, which enables greater representation and learning capabilities in the future dense blocks of the DenseNet architecture.

3.2. Crop Recommendation System

A crop suggestion system is an essential component of contemporary agriculture since it gives farmers access to helpful information and direction about the best possible crop choices. The Random Forest method is a well-known example of machine learning software, and it may be successfully used in the development of such systems. Using the Random Forest algorithm, we will investigate the notion of a crop recommendation system. The capacity of the Random Forest algorithm to handle complicated and non-linear connections in the data is one of its strengths. It is able to capture interactions between a wide variety of agricultural elements and determine the characteristics that are most significant for crop recommendation.

In addition, in comparison to individual decision trees, Random Forest is less likely to result in overfitting, which in turn leads to improved generalization capabilities. An ensemble learning method called the Random Forest technique may be used for both classification and regression. It is a mixture of decision trees, where many trees are built, and the forecasts from each tree are combined to provide a single overall forecast. The random forest approach creates many decision trees, commonly known as "forests," during the training phase and selects a subset of attributes at each split point in the decision tree. This is in contrast to a standard decision tree approach, which would use all of the data available. A method known as bootstrap aggregating or bagging is used in order to accomplish this goal. This method contributes to the reduction of variation as well as overfitting that may take place in decision trees. Figure 2 illustrates the Random Forest algorithm.

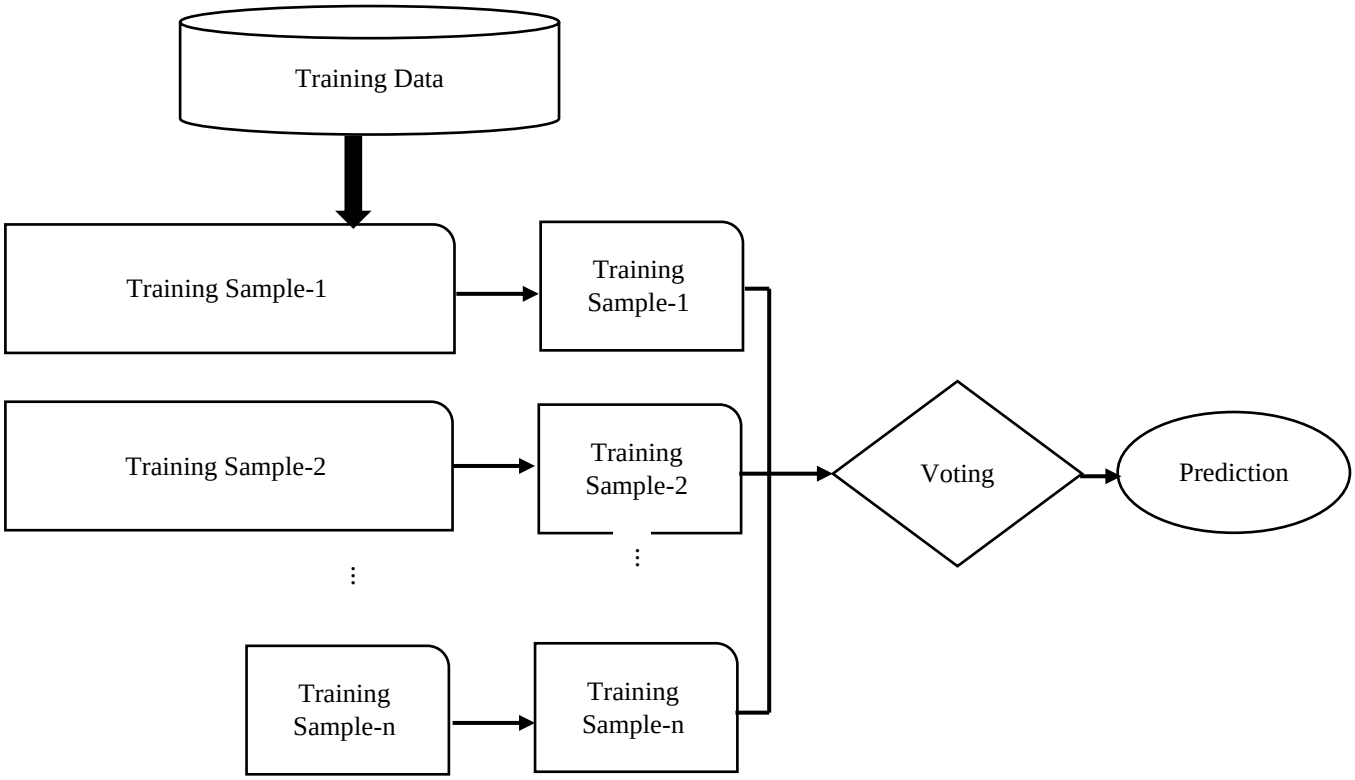


Fig. 2 Random forest classifier

Algorithm: Random Forest based crop recommendation system

Input: A dataset consisting of agricultural features such as soil type, climate conditions, rainfall, temperature, nutrient levels, historical crop yields, and successes.

Step 1: Divide the dataset into a training set and a testing set. The training set will be used to train the Random Forest model, while the testing set will be used to evaluate its performance.

Step 2: Random subsets of the training data are to be selected with replacement (bootstrapping) to create multiple training datasets of the same size as the original dataset.

Step 3: For each subset, a decision tree will be constructed using a modified version of the random forest algorithm. At each node, the best split will be selected based on a feature subset randomly chosen from the available features.

Step 4: A specified number of decision trees will be created, each trained on a different bootstrap subset of the training data.

Step 5: When a new input sample consisting of agricultural parameters (e.g., soil type, climate conditions) is provided, it will be passed through each decision tree in the ensemble. The predictions from each tree will be obtained.

Step 6: A voting mechanism (e.g., majority voting) will be used to determine the final crop recommendation based on the predictions from the ensemble of decision trees.

Step 7: The performance of the Random Forest model will be evaluated using the testing set. Metrics such as accuracy, precision, recall, or F1 score will be calculated to assess the model's effectiveness in providing accurate crop recommendations.

A random forest classifier is an ensemble of decision trees, where each tree is trained on a different subset of the data, and the final prediction is based on the majority vote of the predictions made by all of the trees in the forest. A random forest classifier may be thought of as a voting system for classification problems. In other words, the result is the culmination of the combined predictions made by all of the trees in the forest.

Each tree in the forest is trained using its own unique random subset of the data, and each split point in the forest utilizes a new subset of the characteristics. Because of this unpredictability, the connection between the different trees in the forest is weakened, which contributes to an improvement in the overall performance of the ensemble. In addition to its ease of application and capacity to simultaneously manage a huge number of categorical variables and features, the Random Forest approach is widely regarded as one of the methods that provide the highest level of accuracy and reliability overall. In addition to this, it has a lower risk of producing overfitting than a single decision tree would have, which is a significant benefit. Tuning the hyperparameters is

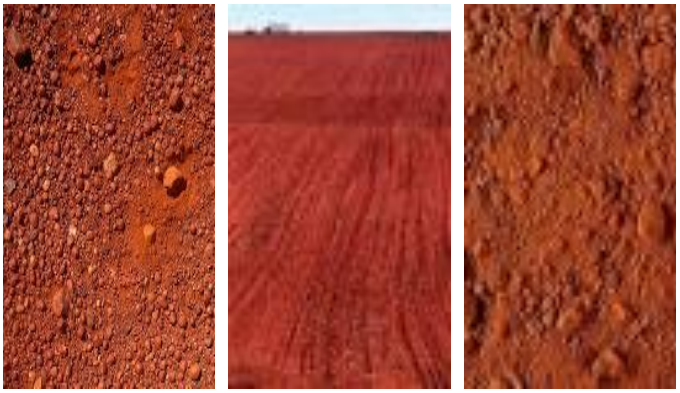
one of the most important steps in the process of optimizing the performance of a Random Forest model. The n_estimators, max_features, and max_depth hyperparameters are three essential hyperparameters that have a substantial influence on the Random Forest method.

- n_estimators is responsible for determining the total number of decision trees that make up the Random Forest. Increasing the total number of estimators may result in a model that is both more reliable and accurate. Nevertheless, this results in an increase in the computing complexity. It is essential to strike an optimal balance between the total number of estimators and the available computing resources.
- max_features hyperparameter is used to specify the maximum amount of features that may be randomly chosen at each split point. It manages the uniqueness and unpredictability of each tree in the forest. Overfitting may be reduced by examining a more manageable subset of features if the max_features parameter is set to a lower value.
- max_depth is a hyperparameter that limits the maximum level of depth that may be reached in any given Random Forest decision tree. A short tree with a low max_depth has a lower propensity to overfit the data, but it may not be as good at making predictions. On the other side, a deep tree with a high max_depth can recognize more intricate patterns, but it also has the potential to overfit the training data.

4. Experimental Results

The findings of the experiments that were conducted to test the suggested model are outlined in this section. The explanation is broken down into two parts: first, the categorization of soils via the suggested architecture of DenseNet201, and second, the selection of crops through the proposed Random Forest Classifier. The compilation includes photographs of five different types of soil, all of which are stored separately inside their directories. The collection contains information on five distinct types of soil, including black soil, cinder soil, yellow soil, clay soil, and laterite soil. Figure 3 displays some of the sample photographs included in the dataset. It is necessary to make certain adjustments to the model in order to use the DenseNet-201 architecture, which was first developed for picture classification for the classification of soil. Figure 1 illustrates the alterations that correlate to the proposed DenseNet-201 model, and section III discusses those very same modifications. The process may be divided up into many distinct steps. To begin, a dataset is produced by collecting photographs of soil together with any relevant information about the soil. A total of five unique soil types should be included in the dataset in order to ensure that the model is able to classify a wide range of soils accurately. The success of the model is significantly dependent on a number of criteria, one of which is the quality of the dataset as well as its variety.

(a) Yellow Soil



(b) Laterite Soil

(c) Clay Soil

(d) Cinder Soil

(e) Black Soil

Fig. 3 Soil sample images in the dataset

The next thing that has to be done is the "pre-processing" of the photos that are part of the dataset. In many instances, this necessitates both the scaling of pictures to a consistent dimension and the standardization of the pixel values. The usage of data augmentation techniques is another way that may be used to attain the goal of increasing the variety of the dataset approaches, as cropping, flipping, and random rotations may be used in order to improve the dataset and make the model more relevant to a larger variety of scenarios. These improvements can be made by using these approaches.After the pre-processing step, the dataset is partitioned into a training set and a validation set so that the recommended model may be trained on it. The performance of the updated DenseNet-201 model is evaluated with the use of a loss function such as categorical cross-entropy. The model is then supplied with training data. During the process of backpropagation, an optimizer such as Adam or SGD is used to make adjustments to the model's weights in response to the estimated gradients.

Throughout the training phase, it is essential to continually assess the performance of the model on the validation set in order to keep track of its evolution. A number of other performance metrics, including recall, accuracy, and precision, in addition to the F1 score, may be used in order to evaluate the effectiveness of the model. It is possible to improve the performance of the model by adjusting hyperparameters such as the learning rate and the batch size. If the first findings from the DenseNet-201 model are unacceptable, then the model as a whole may be modified.

In order to do this, you will need to defrost each layer of the model before carrying on with the training. When there is a large collection of soil imaging data accessible, fine-tuning is an extremely beneficial process to use.Following training, the final model is evaluated using a distinct testing set that it has neither seen nor been exposed to during validation or training. This provides a reasonable assessment of the model's efficacy as well as its potential to generalize to unknown soil samples.

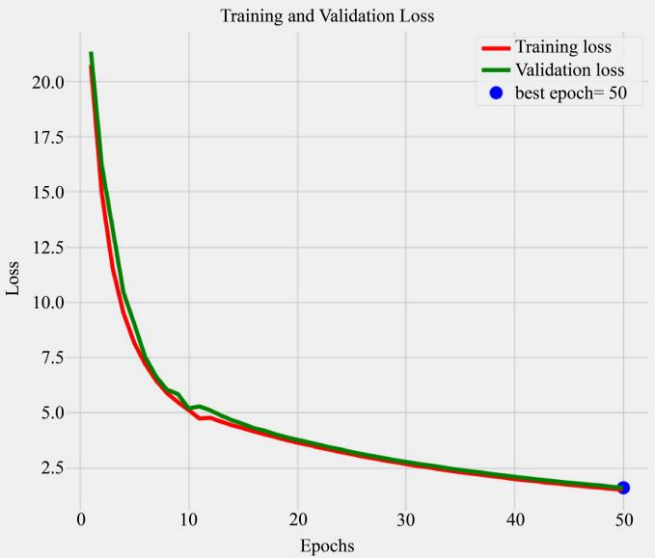


Fig. 4 Training and validation loss plot of the proposed DensNet201 model

Fig. 5 Training and validation accuracy of the proposed DensNet201 model

A DenseNet-201 model's performance during training is shown in Figure 4. The loss of the model over time, which is a measure of how well the model fits the training set of data, is shown on the training plot. A new validation set of data is shown in the validation plot to show the model's loss over time. The mathematical representation of Training and validation loss is shown in Equations (1) and (2), respectively.

$$Training\ Loss\ (L\_train) = (1/N) * \Sigma(y\_pred - y\_true)^2 \quad (1)$$

$$Validation\ Loss\ (L\_val) = (1/M) * \Sigma(y\_pred\_val - y\_true\_val)^2 \quad (2)$$

Where N is the number of training samples, M is the number of validation samples, y_pred is the predicted output, and y_true is the true output. The loss will steadily decrease in a perfect training plot, showing that the model is becoming better at matching the training data. A good validation plot will have a pattern that is comparable to the training loss, even if the loss may not decrease as quickly. This is because the validation set is more challenging than the training set, which raises the possibility that the model will not be able to fit it as perfectly.

Figure 5 shows the training and validation accuracy charts, which provide evidence of the proposed DenseNet-201 model's performance during training. The training plot shows how the model's accuracy increases over time, showing that it can successfully classify the training data. The validation plot, on the other hand, illustrates the model's accuracy using a distinct validation dataset that was not included in the training phase. The training and validation accuracy is represented in Equations (3) and (4) respectively.

$$Training\ Accuracy = (Number\ of\ correctly\ classified\ training\ samples) / (Total\ number\ of\ training\ samples) * 100\% \quad (3)$$

$$Validation\ Accuracy = (Number\ of\ correctly\ classified\ validation\ samples) / (Total\ number\ of\ validation\ samples) * 100\% \quad (4)$$

Ideally, a training plot should exhibit increasing accuracy over time, demonstrating the model's improvement in categorizing the training data. While the validation accuracy may not increase as rapidly as the training accuracy, a good validation plot will show a similar pattern. This is because the validation set is typically more challenging to categorize accurately than the training set. Overfitting becomes apparent when the training accuracy continues to rise while the validation accuracy plateaus or declines. Overfitting occurs when the model excessively memorizes specific instances from the training set, indicating that it has learned the training data too well.

This can happen when the training set is too small or when the model is overly complex. A well-trained model often exhibits an increasing training plot and a moderately increasing validation plot, albeit at a slower pace. Divergent patterns between the training and validation plots indicate that the model is underperforming and may require adjustments. Additional factors to take into account while analyzing training and validation accuracy charts include the following: It is important to take into consideration the number of epochs, which is the number of times that the model comes into contact with the full training set. When training a model for too few epochs, it is possible that the model will not be able to learn from the training data properly; nevertheless, training a model for too many epochs might lead to overfitting.

The size of the training set is an essential component to consider. Models that have been trained on bigger training sets are more likely to generalize efficiently, while models that have been trained on smaller sets may have difficulty generalizing to data they have not seen before. It is important to take into account how complicated the model is. While a more complicated model may be able to attain better accuracy on the training data, it also runs a larger risk of overfitting the data. The accuracy of the proposed model is compared with other models, such as conventional DensNet, VGG and ResNet. The corresponding comparison results are reported in Table 1. Crop selection comes after soil classification is finished using the proposed DenseNet-201 model. Here, crops are chosen using a Random Forest classifier. Crop selection is the process of choosing the plant species that will produce the greatest harvest in a region, depending on the characteristics of the soil in that region.

Table 1. Comparison results of soil classification using the proposed DenseNet201 model

Model Name	Accuracy
AlexNet	85%
VGG	90%
CNN	88%
Conventional DenseNet	94%
Proposed DenseNet 201	**99%**

Table 2. The Accuracy results of the proposed Random Forest classifier

Crop	Precision	Recall	F1-score
Apple	1	1	1
Banana	1	1	1
Blackgram	1	1	1
Chickpea	1	1	1
Coconut	1	1	1
Coffee	1	1	1
Cotton	1	1	1
Grapes	1	1	1
Jute	0.88	0.93	0.90
Kidney beans	1	1	1
Lentil	1	1	1
Maize	1	1	1
Mango	1	1	1
Mothbeans	1	1	1
Mungbean	1	1	1
Muskmelon	1	1	1
Orange	1	1	1
Papaya	1	1	1
Pigeon peas	1	1	1
Pomegranate	1	1	1
Rice	0.95	0.91	0.93
Watermelon	1	1	1
Accuracy = 0.99			

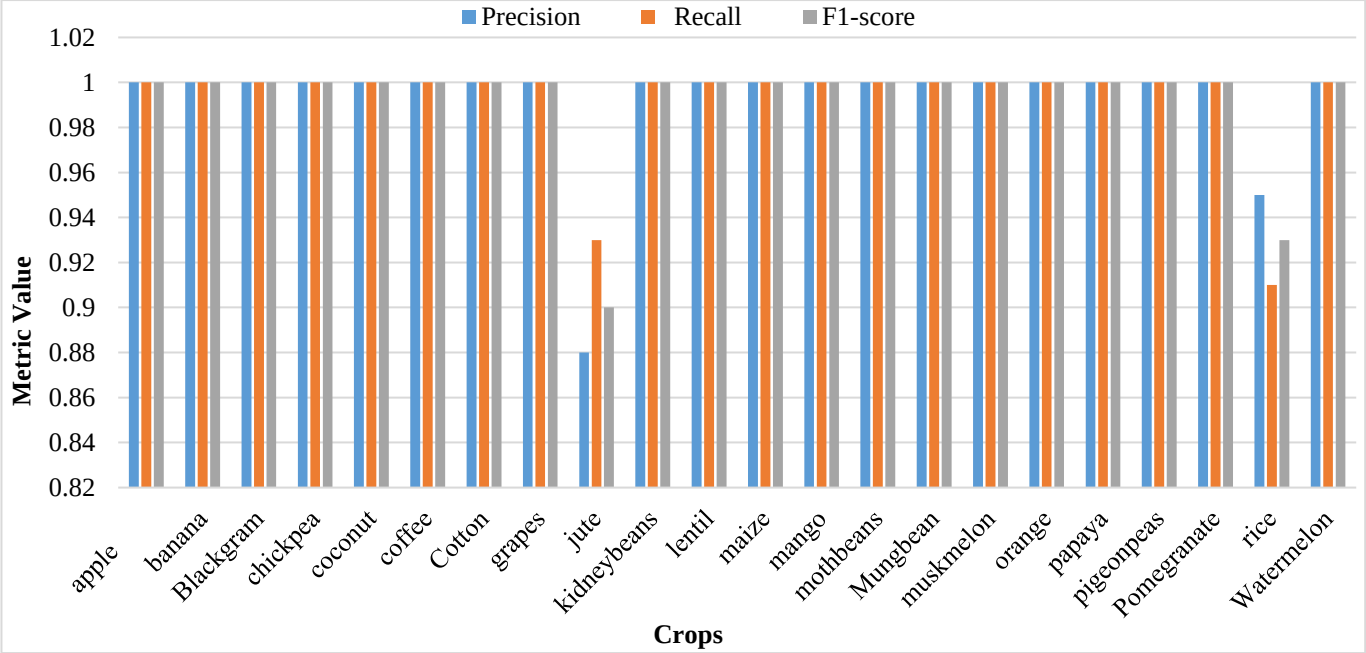


Fig. 6 Performance bar graph results of proposed Random Forest classifier

Information on the various kinds of soil in the region is provided by the DenseNet-201 model, which was trained on soil categorization. The Random Forest classifier is used to forecast which crops would perform best in each kind of soil once the proposed approach has identified the various soil types. A technique for machine learning called Random Forest is often used for classification problems. It works by constructing a number of decision trees and then making a forecast using a voting system. The Random Forest classifier considers a number of variables, including rainfall, temperature, pH level, humidity, NPK metrics, and other significant properties, such as soil type, that are crucial for crop development. These criteria enable the classifier to categorize the various soil types into separate groups and suggest crops that are appropriate for each category. The accuracy of the proposed Random Forest classifier is reported in Table 2. The performance of the proposed Random Forest classifier is shown in Figure 6. The 99% Accuracy is achieved by using a Random Forest classifier for crop selection, which can result in a reliable and accurate technique. This is because it uses the information obtained by the DenseNet-201 model's soil classification to make informed decisions about crop suitability. This combined approach helps to optimize agricultural practices by matching the soil characteristics with the right crops, which ultimately leads to higher yields and more efficient use of resources.

5. Conclusion

The combination of AI and IoT technologies in the agricultural sector has resulted in a revolution in the categorization of soils and the algorithms that propose crops. These developments make it possible for farmers to make educated choices based on data that is both accurate and current. Farmers are able to improve resource use, increase crop output, and encourage sustainable farming practices if they use internet-of-things devices to gather data on soil conditions and climatic parameters and then apply artificial intelligence algorithms to evaluate this data. Utilizing the capabilities of AI and the IoT to develop intelligent agricultural systems that optimize efficiency and output while reducing environmental effects is the future of agriculture.

The proposed modified DenseNet model has proved to be a compelling choice for analyzing and classifying soil images, aiding in better understanding and management of soil properties in agricultural and environmental applications. Random Forest offers robust predictions, handles high-dimensional data, is flexible and non-parametric, is robust to overfitting and noisy data, provides interpretability, and is scalable. These advantages make it a suitable choice for building a crop recommendation system that can assist farmers in making informed decisions, optimizing crop selection, and improving agricultural productivity.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

Author 1: Draft manuscript preparation, study conception and Design, data collection.
Author 2: Study conception and design, data collection.
Author 3: Study conception and design, data collection.

References

[1] A. Subeesh, and C.R. Mehta, "Automation and Digitization of Agriculture Using Artificial Intelligence and Internet of Things," *Artificial Intelligence in Agriculture*, vol. 5, pp. 278-291, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Garima Singh, Anamika Singh, and Gurjit Kaur, "Role of Artificial Intelligence and the Internet of Things in Agriculture," *Artificial Intelligence to Solve Pervasive Internet of Things Issues*, pp. 317-330, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] T. Saranya et al., "A Comparative Study of Deep Learning and Internet of Things for Precision Agriculture," *Engineering Applications of Artificial Intelligence*, vol. 122, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mohd Javaid et al., "Understanding the Potential Applications of Artificial Intelligence in Agriculture Sector," *Advanced Agrochem*, vol. 2, no. 1, pp. 15-30, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Elsayed Said Mohamed et al., "Smart Farming for Improving Agricultural Management," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 971-981, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yaganteeswarudu Akkem, Saroj Kumar Biswas, and Aruna Varanasi, "Smart Farming Using Artificial Intelligence: A Review," *Engineering Applications of Artificial Intelligence*, vol. 120, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Bam Bahadur Sinha, and R. Dhanalakshmi, "Recent Advancements and Challenges of Internet of Things in Smart Agriculture: A Survey," *Future Generation Computer Systems*, vol. 126, pp. 169-184, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Puneet Kumar Aggarwal et al., "AIIoT for Development of Test Standards for Agricultural Technology," *Intelligence of Things: AI-IoT Based Critical-Applications and Innovations*, pp. 77-99, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Christine Dewi, and Rung-Ching Chen, "Decision Making Based on IoT Data Collection for Precision Agriculture," *Intelligent Information and Database Systems: Recent Developments*, pp. 31-42, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Gulbir Singh, and Kuldeep Kumar Yogi, "Internet of Things-Based Devices/Robots in Agriculture 4.0," *Sustainable Communication Networks and Application: Proceedings of ICSCN*, pp. 87-102, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Tien-Heng Hsieh, and Jean-Fu Kiang, "Comparison of CNN Algorithms on Hyperspectral Image Classification in Agricultural Lands," *Sensors*, vol. 20, no. 6, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Zhenrong Du et al., "Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method," *Remote Sensing*, vol. 11, no. 7, pp. 1-21, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Vittorio Mazzia, Aleem Khaliq, and Marcello Chiab, "Improvement in Land Cover and Crop Classification Based on Temporal Features Learning From Sentinel-2 Data using Recurrent-Convolutional Neural Network (R-CNN)," *Applied Sciences*, vol. 10, no. 1, pp. 1-23, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ce Zhang et al., "Scale Sequence Joint Deep Learning (SS-JDL) for Land Use and Land Cover Classification," *Remote Sensing of Environment*, vol. 237, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Rahim Azadnia et al., "Developing an Automated Monitoring System for Fast and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Network and Machine Vision System," *Measurement*, vol. 190, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Pallavi Srivastava, Aasheesh Shukla, and Atul Bansal, "A Comprehensive Review on Soil Classification using Deep Learning and Computer Vision Techniques," *Multimedia Tools and Applications*, vol. 80, pp. 14887-14914, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Amit Bhola, and Prabhat Kumar, "Performance Evaluation of Different Machine Learning Models in Crop Selection," *Robotics, Control and Computer Vision*, pp. 207-217, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Liheng Zhong, Lina Hu, and Hang Zhou, "Deep Learning Based Multi-Temporal Crop Classification," *Remote Sensing of Environment*, vol. 221, pp. 430-443, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Zeel Doshi et al., "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] Andreas Kamilaris, and Francesc X. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70-90, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Krupa Patel, and Hiren B. Patel, "A State-of-the-Art Survey on Recommendation System and Prospective Extensions," *Computers and Electronics in Agriculture*, vol. 178, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] P. Parameswari, and C. Tharani, "Crop Specific Cultivation Recommendation System Using Deep Learning," *Information and Communication Technology for Competitive Strategies*, pp. 781-787, 2023. [CrossRef] [Google Scholar] [Publisher Link]