International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Comprehensive Review on Optimizing Agricultural Production Using Machine Learning and IoT

Anuradha Brijwal[1*], Praveena Chaturvedi[2]

Submitted:03/02/2024 **Revised**: 11/03/2024 **Accepted**: 18/03/2024

Abstract "Optimize Agriculture Productionusing Internet of Things and ML is a rapidly expanding field in agricultural. Crop prediction holds utmost significance in production. The smart information system assists farmers by providing information relating to all environmental factors, suggestions and offer of crop sowing recommendation. Generally, Farmers choose their crops without taking the environment into account. Poor harvest results from it. These are the concerns that farmers and agriculturists are currently facing. These are the current issues of the agriculturists and farmers.Machine learningtechniques and IOT offer a promising solution by automating crop recommendations. This study reviews the production of crop using machine-learning technique and IOT. The suggested system makes accurate predictions about which crops would be most suited for a given site by utilising a number of features, such as soil and weather data.The potential for such a novel method to transform crop recommendation might help farmers to increasing crop production.With the help historical dataset, we trained and tested the ML algorithms with different parameters, ultimately achieving near-perfect accuracy. All models exhibit accuracy levels exceeding 94% on a consistent basis, with the best accuracy yet measured reaching an astounding 99.7%.This study presents perfectly accurate machine-learning models for crop recommendation. The method accurately predicts the most suited crops by utilising a variety of characteristics, including soil and weather data. This technology has the potential to be revolutionary in that it can improve agricultural yields, sustainability, and overall profitability, which will help farmers of all sizes. For higher production we have to move from traditional approach to advanced approach. We are convinced that with the help of latest approach, change crop recommendations and help guarantee a long-term. With more thaneight billion people on the planet, our dependence on agriculture for food necessitates the establishment of resilient and sustainable agricultural systems. The manuscript's future prospects include utilising our models to develop an end-to-end system and surveying farmers to obtain numerical estimates of the impacts.

Key Terms*—ML algorithm, Data Analysis, Big Data, Crop Recommendation.*

I. Introduction

ML [1] [2] is a rapidly advancing area that empowers computers to train from data without explicit programming, as defined by Arthur Samuel in 1959. By training on large datasets, ML algorithms [3] are capable of producingwell-informed judgements.

ML algorithms [3] are capable of producing well-informed judgements.A vital industry in the world, agriculture is largely dependent on farmers' capacity to produce crops that are both profitable and sustainable. The improper crop choice can have a big impact on agricultural production, which can result in lower productivity and even financial losses. Ignoring critical elements such as soil quality, market demand, and climate compatibility can prevent crops from growing to their maximum potential and thriving. Unfavourable crop choices may result in inadequate climatic adaptation, which may impair growth, make plants more susceptible to diseases and pests, and reduce output as a whole. Furthermore, yield misaligned with market demand might struggle to find buyers or fetch favourable prices, further burdening farmers economically. To overcome these challenges and optimize crop yield for long-term agricultural viability, ML-based crop recommendation systems save the day by giving farmers the information they need to make wise choices.

How farmers understand and improve their operations is being revolutionised by the intersection of ML and agricultural data. The increasing availability of data from different sources has made machine learning (ML) techniquescapable of analysing massive volumes of

*1*Research Scholar, Department of Computer Science, kanya Gurukul Campus Dehradun, Gurukul Kangri (Deemed to be University), Haridwar, Uttarakhand, India*
2Professor Department of Computer Science, kanya Gurukul Campus Dehradun, Gurukul Kangri (Deemed to be University), Haridwar, Uttarakhand, India
*1*21523002@gkv.ac.in,*
2praveena.chaturvedi@gkv.ac.in
*1*https://orcid.org/0009-0005-7485-6172*

data and producing informative results.These techniquesenable farmers to make data-driven decisions on crop selectionand yield prediction by revealing intricate patterns, correlations, and predictive models concealed within the data. The agriculture sector will ultimately see increased productivity and profitability as a result of this integration, which improves efficiency, resource optimisation, and sustainable practices.

Crop prediction systems use information from multiple sources of data,like market, soil, and climate data. These platforms offer optimal techniques for growing specific crops and make AI procedures to conjecture which harvests will flourish in given regions.

ML-based crop prediction systems could increase the sustainability and productivity of agriculture. By guiding harvesterin choosing suitable crops, these systems can increase crop yields and reduce resource consumption. They also make agriculture more resilient to climate change[4]. Moreover, ML has the potential to tackle a plethora of additional agricultural difficulties [5], including but not limited to crop yield prediction, pest and disease identification, crop production optimisation, water efficiency enhancement, and reduced usage of fertiliser and pesticides.

A substantial portion of the planet's food and fibre supply comes from crops. As the global population approaches more than ten billion by 2050, the World Resource Institute strives to address the challenge of sustainably feeding this growing population. Thus, it becomes essential to achieve a high-quality crop output. agricultural choice has a significant impact on agricultural yields and profitability. predicting crop performance based on location is becoming more difficult due to weather change and other environmental factors.

In this paper, we use ML-algorithm to predict crop and provide information to harvester. We begin by collecting and preprocessing the dataset. Next, we train and test our models using a range of characteristics, such as soil and climate conditions. To find out if the model works better when using a mix of multiple parameters, we also investigate parameters engineering principles and add them as additional parameters to the dataset. Additionally, we comprehensively highlight general challenges in agriculture, particularly within the use of ML.

II. Background Survey

A. Machine Learning:

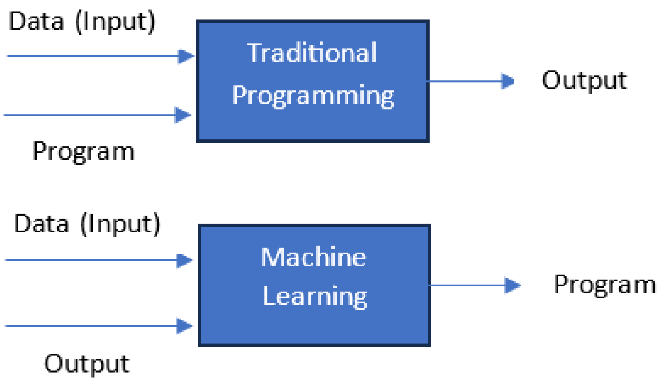

Fig. 1: Traditional Programming Vs. Machine Learning

Machine learning is a field of study that enables computers to learn from data without being explicitly programmed. This process involves transforming data into numerical representations and identifying patterns within certain figures. The patterns found aid in forecasting results for fresh data points. Unlike traditional programming, where code defines the steps to solve a problem, In ML-algorithm, a model is trained using dataset, allowing it to learn and solve problems autonomously is shown in Figure 1. Based on how machines learn, there are three different types of ML-algorithms:

1. **Supervised Learning**: In supervised machine learning, models are trained on labelled data. Next, the model gains the ability to forecast results for fresh, unlabelled data. Examples of supervised learning algorithms include decision trees, logistic regression, support vector machines, and neural networks.

2. **Unsupervised Learning**: Unsupervised learning involves training a model on a set of unlabelled data. Without labels, the model learns to identify trends and cluster related data pieces.

3. **Reinforcement Learning**: Reinforcement learning is a type of ML algorithm that enables software agents and machines to automatically evaluate the optimal behaviour in a particular context or environment to improve its efficiency, i.e., an environment-driven approach. Reinforcement learning is based on reward or penalty, and its ultimate goal is to use insights obtained from environmental activists to take action to increase the reward or minimize the risk.

B. Machine Learning Algorithms Used:

In this survey, we focus on specific machine learning algorithms used in the study:

1. K-Nearest Neighbours (KNN): A supervised learning technique that uses the labels of the k neighbours in the training set that are most similar to predict labels for new data.

2. Naive Bayes (NB): A supervised learning technique that makes it easy to train and interpret by assuming feature independence and applying the Bayes theorem.

3. Random Forest (RF): An ensemble learning algorithm consisting of multiple decision trees trained on different subsets of the data, and the final prediction is made through majority voting.

4. Logistic Regression (LR): A statistical technique that estimates the likelihood that an event will occur. It simulates how category dependent characteristics and one or more independent features relate to one other.

5. Decision Tree (DT): A hierarchical structure that represents a supervised learning algorithm, where nodes show decisions and branches show possible outcomes based on input data.

6. Support Vector Machine (SVM): A supervised learning algorithm that finds a hyperplane to separate data into two classes.

7. Neural Network: A neural network, which is modelled after the structure of the human brain, is made up of layers of interconnected neurons, or nodes. It learns by adjusting weights and biases during training, and the output is calculated using activation functions.

These machine learning algorithms play a pivotal role in the development of crop recommendation systems and provide valuable insights to farmers for making informed decisions in agriculture.

C. Frameworks

Multiple frameworks can help with large-scale dataset analysis. Even though each has particular advantages and disadvantages, they can also complement each other in certain scenarios.

Machine learning models excel at learning patterns from large datasets and making predictions or decisions autonomously without human intervention. They are particularly useful when accuracy and predictive capabilities are essential.

On the other hand, big data processing frameworks like MapReduce are designed to efficiently process vast amounts of data quickly. Large graph processing, analytics, data mining, and data warehousing are among the typical uses for them. These frameworks are optimized for handling large-scale data processing efficiently.

ML-algorithm and big data frameworks can work together synergistically. For instance, ML-Model can be trained using ML-algorithm and then utilized by big data frameworks for predictions. This combination leverages the strengths of both approaches, enabling accurate predictions while efficiently handling massive datasets.

When choosing between machine learning and big data processing frameworks, the decision often depends on the specific needs of the task at hand. If accuracy, predictive capabilities, and direct predictions from data are crucial, machine learning may be the preferred option. On the other hand, if the main focus is on processing massive datasets quickly and efficiently for data-centric tasks, big data processing frameworks can be the more suitable choice.

Ultimately, the ideal tool depends on the project's requirements, and in some cases, a combination of ML and big data frameworks can provide a comprehensive solution to address accuracy, speed, scalability, and ease of use.

D. Existing Research in Crop Recommendation

Existing research in the field of crop recommendation has seen some growth in recent years [22]. Several studies and systems have been developed to address the challenges of crop recommendation and help farmers make knowledgeable crop selection decisions.

1. Priyadharshini A et al. (2021) introduced an "Intelligent Crop Recommendation System" [23].

2. Zeel Doshi et al. (2018) presented a system called "AgroConsultant" [24].

3. SM Pande et al. (2021) proposed a user-friendly yield prediction system for farmers in their paper [25].

4. RK Rajak et al. (2017) proposed a model using a majority voting technique with SVM and ANN to recommend crops [26].

5. Reddy et al. conducted a survey of existing techniques for crop recommendation in their paper [27].

6. Ghadge et al. presented a theoretical approach to crop recommendation in their paper [28].

7. Kulkarni et al. worked on improving crop productivity through a crop recommendation system using assembling techniques [29].

8. Pudumalar et al. presented a highly cited paper on a similar approach using machine learning for crop recommendation in Tamil Nadu, India [30].

9. Konstantinos G. Liakos et al. conducted a review on "Machine Learning in Agriculture" covering various applications but did not mention crop recommendation [31].

10. Ayaz Muhammad et al. focused on the use of the Internet of Things and sensors for agricultural data collection [32].

Despite some research in this area, there remains a relatively limited amount of literature on crop recommendation compared to other agricultural-related topics. The difficulties associated with bringing machine learning to agricultural, as well as the fundamental obstacles facing the agriculture sector, may contribute to this scarcity [22].

III. Our Contribution

As technology advances and more data becomes available, it is likely that research in crop recommendation and related fields will continue to grow, helping farmers optimize crop selection and achieve better productivity in their agricultural practices.

Our paper addresses several limitations found in the existing literature on crop recommendation models (covered in section II-D). Many previous works lack comprehensive overviews of their research process, including dataset sources, model accuracy, and training and testing details. Additionally, implementation details and specified features are often missing, and Certain studies solely offer surveys or theoretical analysis on themes related to crop prediction.

To overcome these boundaries, our contribution involves the development of comprehensive crop predictionmodels. Every stage of our procedure, from information collection and engineering to model training and assessment. Our method outperforms all crop predictionmodels in the literature currently in circulation in terms of accuracy. Feature engineering plays a crucial role in enhancing the data's utility for ML, and In order to maximise accuracy, I carefully evaluate the data using seven distinct machine learning algorithms in different configurations.

The key aspects of our contribution are as follows:

1. Comprehensive Crop Recommendation System: I present a detailed and crop recommendation system, addressing the limitations identified in previous works.

2. Training with Various Algorithms: We train our models using several machine learning algorithms and explore different setups for every model to improve accuracy and performance.

The algorithms we employ are:

• LR

• DT

• RF

• KNN

• NB

• SVM

• NN

3. Addressing Challenges in Agriculture: We draw attention to difficulties faced within the agricultural industry, both generally and specifically when usingML techniques to agricultural data.

4. Future Work: I propose several valuable concepts for future research in our field, detailed in section VII of the manuscript.

Overall, our research makes an important addition to the crop prediction field. Withthorough methodology, superior precision, and focus on characteristic engineering represent innovative inputs. I think farmers, agricultural researchers, and other farming industry stakeholders will find our findings useful, helping them make informed decisions and optimize crop productivity.

IV. Overview Of Data, Methodology, And Experimentation

A. Overview of Data:

For our model, I preprocess a Kaggle dataset that already exists [33]. A sample of the data's first few rows is shown in Table I. The visual illustration that shows the characteristics and their number can be observed in Figures 4 and 5, respectively. Figure 3 illustrates the pair plot, showing connections among various attributes shown as a matrix.

The dataset is derived from augmenting weather, and soil data available. I used a total of twenty-two unique labels, as shown in Table II. The labels were fetched from a data source of around 50k records and then reduced to approximately 2.2k records, ensuring that each setting had one good crop.

S. No.	N	P	K	TEMP	HUM	ph	rainfall	label
1	89	43	41	19.88	81.00	6.10	203.936	rice
2	71	54	16	22.61	63.69	5.75	87.7595	maize
3	40	72	77	17.02	16.99	7.49	88.5512	chickpea
4	13	60	25	17.14	20.60	5.69	128.257	kidneybeans
5	31	72	17	28.69	49.47	5.83	96.3622	pigeonpeas
6	36	58	25	28.66	59.32	8.40	36.9263	mothbeans
7	40	45	18	30.44	55.21	5.26	30.9201	mothbeans
8	19	35	24	27.11	83.64	6.88	49.1196	mungbean
9	57	67	25	32.35	66.61	7.55	64.5588	blackgram
10	32	76	15	28.05	63.50	7.60	43.358	lentil
11	2	24	38	24.56	91.64	5.92	111.968	pomegranate
12	86	76	54	29.32	80.12	5.93	90.1098	banana
13	23	23	27	34.72	51.43	5.16	97.3126	mango
14	28	122	197	19.89	82.73	5.86	69.6626	grapes
15	80	26	55	24.53	88.99	6.14	49.1162	watermelon
16	109	26	45	28.28	90.39	6.22	21.5899	muskmelon
17	30	122	197	21.38	92.72	5.57	106.142	apple
18	13	5	8	23.85	90.11	7.47	103.923	orange
19	69	60	54	36.32	93.06	6.99	141.174	papaya
20	39	5	31	27.10	93.70	5.55	150.95	coconut
21	104	47	18	23.97	76.98	7.63	90.7562	cotton
22	70	38	35	24.40	79.27	7.01	164.27	jute
23	81	30	31	24.65	51.94	7.03	135.139	coffee

TABLE I: Sample of Dataset

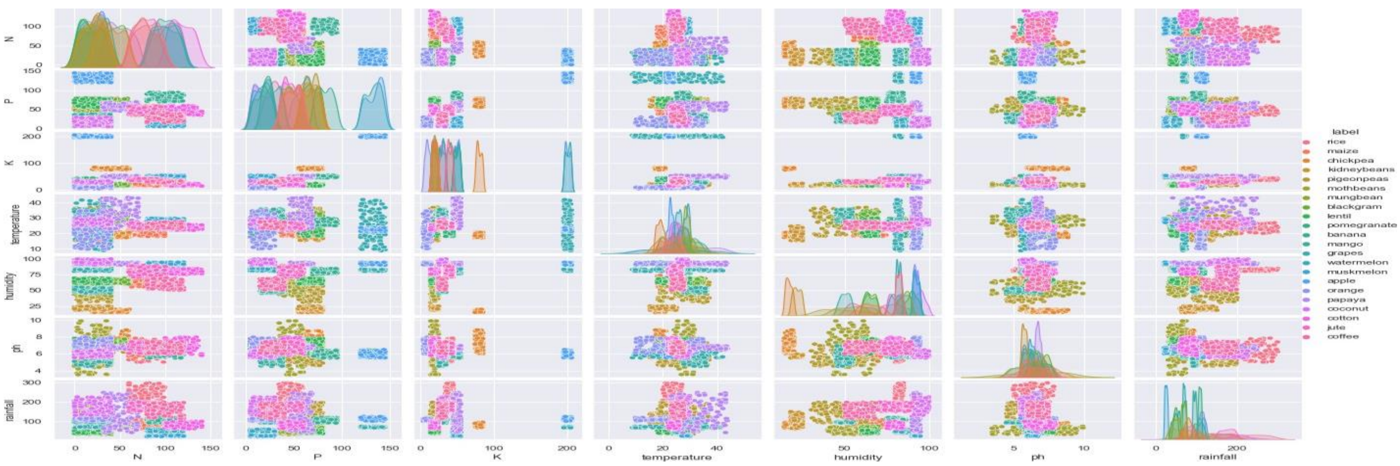

Fig. 2: Pair Plotting of All Data

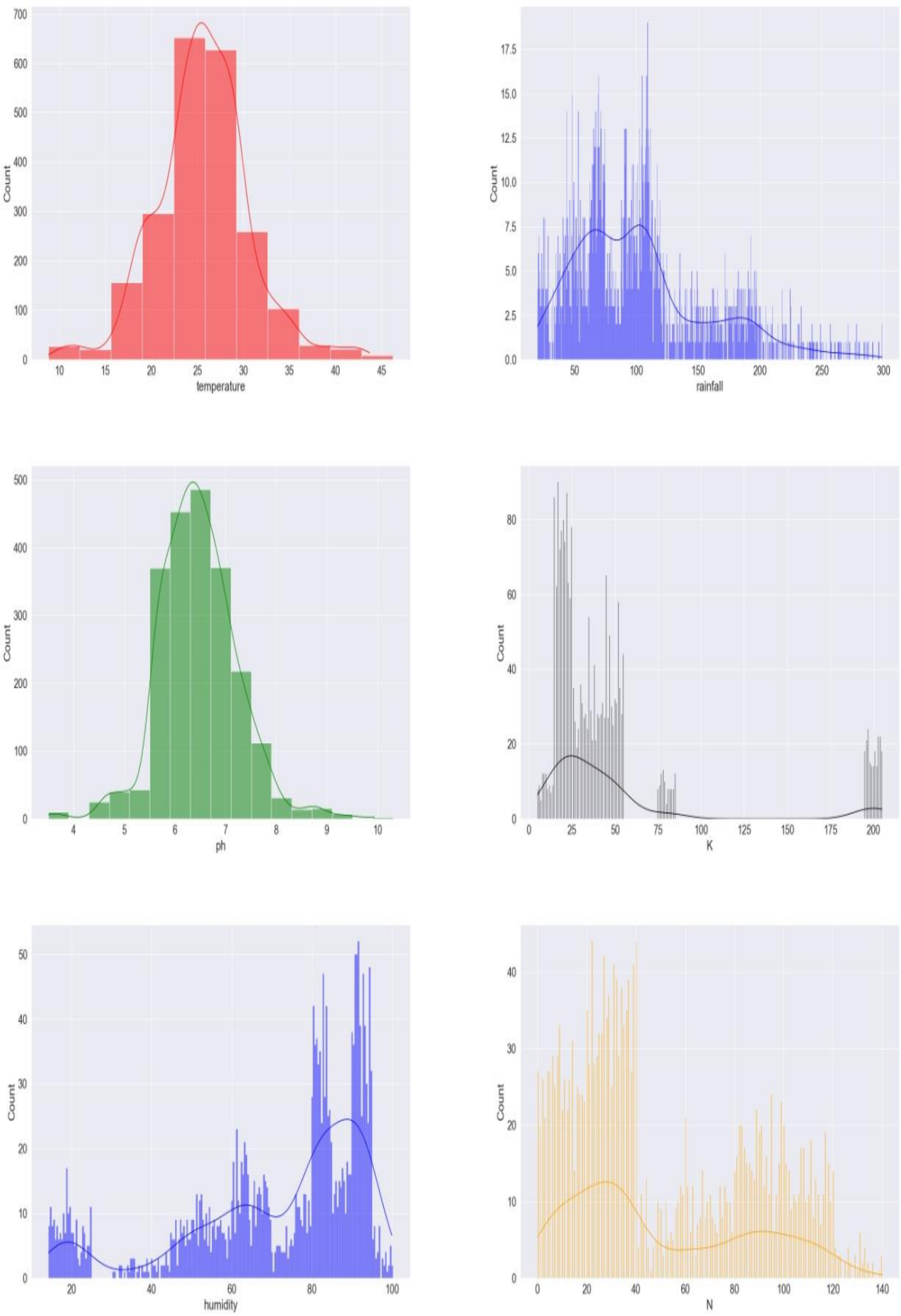

Fig. 3: Feature Graphs

S. No.	***Unique LabelName***
i	apple
ii	banana
iii	blackgram
iv	chickpea

v	coconut
vi	coffee
vii	cotton
viii	grapes
ix	jute
x	kidneybeans
xi	lentil
xii	maize
xiii	mango
xiv	mothbeans
xv	mungbean
xvi	muskmelon
xvii	orange
xviii	papaya
xix	pigeonpeas
xx	pomegranate
xxi	rice
xxii	watermelon

TABLE II: List of all Labels

B. Methodology:

Our methodology, as depicted in Figure 6, outlines the steps we followed to train various models using different machine learning algorithms. We iterated through the following steps for each selected algorithm listed in Section III.

1. Input Data: The input to our system consists of soil and environmental characteristics data. Data quality and quantity play a crucial role in the correctness of the model. I made sure the information was accurate, properly marked, and devoid of anomalies.

2. Initial Processing: I prepared the data for ML techniques by cleaning it up, eliminating outliers, and formatting it. This step involved handling redundant and empty records, segregating characteristics listed in the label section, conducting feature engineering, and plotting and visualising data to look for anomalies.

3. Selection of ML Algorithm: With every cycle, I selected among the seven chosen algorithms. I then proceeded with preprocessing and to fine-tune the model, test or validate it.

4. Model Installations: To improve the efficacy of tests and cross-validation, I experimented with different arrangements, such as epochs, DT depth, activation functions and the quantity of nearest neighbours. Iwas mindful of the model's performance, as some configurations could impact it negatively.

5. Training Models: In this step, the selected ML algorithm learned the sourceinitial-processed data.

6. Testing: I evaluated the accuracy of the model against test data and measured cross-validation accuracy. If the accuracy was unsatisfactory, we returned to the "Model Configuration" step. In some cases, we experimented with feature engineering. If the model's accuracy and performance were good, we proceeded to the "Choose a Machine Learning Algorithm" step to repeat the process with a different algorithm.

By following this methodology and experimenting with various configurations and algorithms, we aimed to develop robust and accurate crop recommendation models.

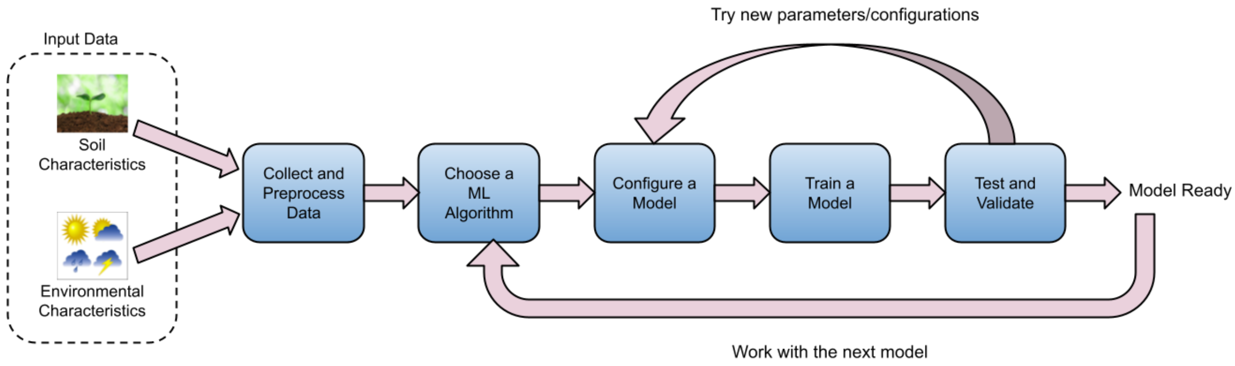

Fig. 4: Methodology

C. Experimentation

In our experimentation, we developed a multi-class neural network model for crop recommendation. Unlike a single-class neural network, which can only classify data into one category, a multi-class neural network can classify data into multiple classes, making it suitable for our crop recommendation task.

The neural network was built using the TensorFlow framework, and we designed a four-layered architecture as shown in Listing 1. The input layer had thirty neurons, followed by twenty neurons in the second layer, ten neurons in the third layer (hidden layers), and twenty-two neurons in the output layer, representing the number of unique crop labels we used for classification.

To optimize the neural network's performance, we experimented with different combinations of activation functions: "relu," "softmax," and "sigmoid." These activation functions help in introducing non-linearity into the model, enabling it to learn complex patterns in the data.

Additionally, we fine-tuned the neural network by experimenting with multiple epoch values. Epochs represent the number of times the entire dataset is passed through the neural network during training. Increasing epoch values can improve the accuracy, but there is a point where it starts overfitting, leading to decreased performance on unseen data.

We used Categorical Cross Entropy as the loss function for training the multi-class classification model. Categorical Cross Entropy measures the distance between the predicted probabilities and the actual labels. Lower values of Categorical Cross Entropy indicate better model performance.

Finally, we utilized the Adam optimizer, an extension of the AdaGrad and RMSProp algorithms, which is effective for a wide range of deep learning problems. The Adam optimizer helps in efficiently updating the model's weights during training, improving convergence and performance.

Listing 1: Neural Network - Multi-Class Crop Recommendation Model

In this code listing, we create a multi-class neural network model using the TensorFlow library for crop recommendation. The neural network architecture consists of four layers, including an input layer, two hidden layers, and an output layer. The model takes input with a shape of (7,) representing the features of the dataset.

```
```python

import tensorflow as tf

# Define the neural network model architecture

model = tf.keras.models.Sequential([

tf.keras.layers.Dense(30, activation='relu', input_shape=(7,)),

tf.keras.layers.Dense(20, activation='relu'),

tf.keras.layers.Dense(10, activation='relu'),

tf.keras.layers.Dense(labels_count, activation='softmax')

])

# Compile the model with appropriate loss function, optimizer, and metrics

model.compile(

    loss=tf.keras.losses.CategoricalCrossentropy(),
```
```

```
    optimizer=tf.keras.optimizers.Adam(),
    metrics=['accuracy']
)
# Train the model using the training data and validate it using the testing data
model.fit(x_train, y_train, epochs=60, validation_data=(x_test, y_test), batch_size=32)
```

Explanation:

1. We import the necessary TensorFlow library to build and train the neural network model.

2. We define the model using `tf.keras.models.Sequential()`, which allows us to stack layers sequentially.

3. The first layer is the input layer with 30 neurons and uses the ReLU (Rectified Linear Activation) function to introduce non-linearity.

4. The second and third layers are hidden layers with 20 and 10 neurons, respectively, and both use the ReLU activation function.

5. The last layer is the output layer with `labels_count` neurons, representing the number of unique crop labels used for classification. It uses the softmax activation function to compute the probabilities of each class.

6. We compile the model using `model.compile()` with the Categorical Crossentropy loss function, suitable for multi-class classification tasks. The Adam optimizer is used to optimize the model's weights, and the accuracy metric is used to evaluate its performance during training.

7. We train the model using `model.fit()` with the training data (`x_train` and `y_train`) for 60 epochs. The validation data (`x_test` and `y_test`) is used to validate the model's performance during training. The `batch_size` parameter determines the number of samples used in each update of the model's weights.

By running this code, the neural network model will be trained and optimized to make accurate crop recommendations based on the provided input features (7 features in this case). The model's performance and accuracy will be displayed during the training process, and it can be further evaluated using test data to assess its generalization capability.

The rest of the models, excluding neural networks, were built and evaluated using different machine learning algorithms. Here are some specific details for each of these models:

1. Decision Tree:- Impurity Measures: For the decision tree algorithm, two impurity measures were used: Gini impurity and entropy. These measures help in determining the quality of a split at each node of the decision tree.

- Max Depth: Another parameter used for the decision tree is the maximum depth of the tree. It controls the depth to which the tree can grow and helps prevent overfitting.

2. K-Nearest Neighbors (KNN):- N Neighbors: The KNN algorithm was experimented with a configuration called "n neighbors." This parameter determines the number of nearest data points from the training set that should be considered when making predictions for a new data point.

3. Support Vector Machine (SVM):-Kernel Configuration: SVMs are based on finding a hyperplane that separates data into different classes. The kernel configuration determines the type of kernel function used in SVMs, such as linear, polynomial, radial basis function (RBF), etc. The choice of kernel function impacts the performance of the SVM.

For all the models, evaluation and cross-validation were performed using the functionality from Listing 2. The cross-validation process helps to assess the model's performance and generalization ability by training and testing it on different subsets of the data.

Overall, these models were experimented with various configurations and hyperparameters to achieve higher accuracy and better performance. The best-performing configurations for each model were determined based on the evaluation results obtained during training and testing.

Please note that specific hyperparameter values and detailed evaluation results are not provided in the given text. These would be included in the full research paper or report that contains the complete experimentation and results section.

Listing 2 provides two functions for model evaluation and cross-validation:

1. `evaluate(my_model)`: This function evaluates the performance of a trained model on the test data. It takes the trained `my_model` as input and uses it to make predictions on the test data (`x_test_data`).

The predictions are compared with the true labels (`y_test_data`) to calculate the accuracy of the model. The function returns the accuracy rounded to three decimal places.

2. `perform_cross_val(my_model)`: This function performs cross-validation on the model `my_model`. Cross-validation is a technique used to assess a model's performance by splitting the data into multiple subsets (folds) and training/testing the model on different combinations of these subsets. It helps to obtain a more robust estimate of the model's performance. The function uses the `cross_val_score` function from the scikit-learn library to perform cross-validation. The `cross_val_score` function takes the model (`my_model`), input features (`features`), and corresponding labels (`labels`) as input, along with the cross-validation method (`cv=kfold`). It returns an array of scores obtained from each fold. The function calculates the mean of these scores and returns the mean cross-validation score rounded to three decimal places.

Both functions are useful for evaluating the accuracy and performance of different models during experimentation and hyperparameter tuning. The results obtained from these functions can be used to compare the performance of different models and configurations to identify the best-performing model for the crop recommendation task.

V. Results And Evaluation

The results and evaluation of the experiments are summarized in Table V. The data was split into 70% training data and 30% testing data for all the models. When using proper configurations, all the models achieved an accuracy of at least 95%. Various configurations were experimented with for each model, and the ones listed in Table V were found to be optimal in terms of performance and accuracy.

One example is the decision tree algorithm, where increasing the depth value led to higher accuracy but also increased the training and prediction time. Using a random forest algorithm with 100 estimators resulted in an accuracy of 99.5%.

For the neural network model, the number of epochs was observed to play a significant role in accuracy and performance. Increasing the number of epochs generally improved accuracy but also increased training time. For instance, an accuracy of 97.73% was achieved with 100 epochs.

S.No.	***Model Name***	***Accuracy %***	***Validation Accuracy%***	***Configurations***	***Precision/Recall***
i	DT	98.091	98.682	Max Depth=10 and with Gini	0.98/0.98
ii	DT	98.091	98.409	Max Depth=10 and with Entropy	0.98/0.98
iii	K-NN	97.936	97.145	n=6	0.97/0.97
iv	LR	94.545	95.955	-	0.94/0.95
v	NB	99.745	99.6	-	0.99/0.99
vi	NN	-	96	Value of epoch=100	0.99/0.99
vii	S-NN	-	96.73	Value of epoch=2000	0.99/0.99
viii	RF	99.245	99.3	n=200	0.99/0.99
ix	SVM	96.827	96.782	kernel=RBF	0.96/0.96
x	SVM	98.342	97.782	kernel=linear	0.98/0.97

TABLE III: Model Accuracy

S.No.	*Label*	*Nitrogen*	*Phosphorous*	*Potassium*	*Temperature*	*Humidity*	*pH*	*Rainfall*
i	apple	21.8	135.22	200.89	23.63	93.33	6.93	113.65
ii	banana	101.23	83.01	51.05	28.38	81.36	6.98	105.63
iii	blackgram	41.02	68.47	20.24	30.97	66.12	8.13	68.88
iv	chickpea	41.09	68.79	80.92	19.87	17.86	8.34	*81.06*
v	coconut	22.98	17.93	31.59	28.41	95.84	6.98	176.69
vi	coffee	102.2	29.74	30.94	26.54	59.87	7.79	159.07
vii	cotton	118.77	47.24	20.56	24.99	80.84	7.91	81.4
viii	grapes	24.18	133.53	201.11	24.85	82.88	7.03	*70.61*
ix	jute	79.4	47.86	40.99	25.96	80.64	7.73	175.79
x	kidneybeans	21.75	68.54	21.05	21.12	22.61	6.75	106.92
xi	lentil	19.77	69.36	20.41	25.51	65.8	7.93	46.68
xii	maize	78.76	49.44	20.79	23.39	66.09	7.25	*85.77*
xiii	mango	21.07	28.18	30.92	32.21	51.16	6.77	95.7
xiv	mothbeans	22.44	49.01	21.23	29.19	54.16	7.83	52.2
xv	mungbean	21.99	48.28	20.87	29.53	86.5	7.72	49.4
xvi	muskmelon	101.32	18.72	51.08	29.66	93.34	7.36	*25.69*
xvii	orange	20.58	17.55	11.01	23.77	93.17	8.02	111.47
xviii	papaya	50.88	60.05	51.04	34.72	93.4	7.74	143.63
xix	pigeonpeas	21.73	68.73	21.29	28.74	49.06	6.79	150.46
xx	pomegranate	19.87	19.75	41.21	22.84	91.13	7.43	*108.53*
xxi	rice	80.89	48.58	40.87	24.69	83.27	7.43	237.18
xxii	watermelon	100.42	18	51.22	26.59	86.16	7.5	51.79

TABLE IV: Features Average Values for Each Crop

For ML Model, recall and precision are imperative measurements for assessing a model's execution. Recall measures the extent of real positive occasions that are accurately recognized by the model, whereas precision measures the extent of positive expectations that are correct. Table III incorporates recall and precision values for each model.

Based on the experimentation, the ModelUsing the Random Forest and Naive Bayesachieved the highest accuracy. However, it was noted that neural systems might perform indeed superior with bigger dataset sizes.

Also, Table IV presents the normal soil and weather characteristics values for each label. This data can be important for agricultural stakeholders and farmers in making choices which crops are appropriate for their particular region's conditions.

The work presented in this study is expected to be helpful for other developers and researchers, providing insights into the effect of distinctive arrangements on the precision and execution of machine learning models for crop prediction.

VI. Challenges In Agriculture

Insects and diseases: Insects and diseases [49] pose a constant threat to agricultural productivity. Invasive species and new strains of diseases can quickly spread and devastate crops, leading to significant economic losses for farmers.

Limited access to technology: In many rural areas, farmers still lack access to modern agricultural technology and tools. This hinders their ability to optimize their practices, make data-driven decisions, and benefit from advancements in agriculture.

Market volatility: The agricultural market can be volatile, with fluctuating prices and demand for crops. Farmers may struggle to predict market trends, making it challenging to plan their production and manage their profits effectively.

Labour deficiencies: Numerous districts confront labour deficiencies in horticulture, as youthful individuals progressively relocate to urban ranges in look of superior openings. The lack of skilled labour can prevent agricultural productivity and efficiency.

High production costs: The cost of inputs such as seeds, fertilizers, and machinery can be substantial for farmers. High production costs can limit their profitability and ability to invest in more efficient and sustainable practices.

Dependency on traditional practices: Some farmers may be resistant to adopting new technologies or practices due to cultural or economic reasons. The reliance on traditional methods can hinder progress and improvements in agricultural productivity.

Lack of data and infrastructure: In many developing regions, there is a lack of robust data on agricultural conditions and practices. Additionally, inadequate infrastructure, such as poor road networks and limited access to markets, can hamper agricultural development.

Sustainable and Resilient Agriculture: In the face of these challenges, promoting sustainable and resilient agriculture becomes crucial. Sustainable practices aim to optimize resource use, reduce environmental impact, and maintain productivity over the long term. Resilient agriculture involves building systems that can adapt to changing conditions, including climate change and market dynamics.

Data analytics and Machine learning can play a crucial part in tending to a few of these challenges. By analysing huge amount of machine learning models, data can give profitable experiences, help predict pest outbreaks, optimize irrigation, and enhance crop recommendation systems to maximize productivity and sustainability. However, overcoming challenges in horticulture requires an all-encompassing approach, combining mechanical headways, arrangement intercessions, and community engagement to construct a more flexible and economical rural division.

Indeed, these are significant challenges that the agriculture sector faces, and they can have a profound impact on farmers' livelihoods and food production. Tending to these challenges requires a combination of technological, policy, and social interventions. Now explore potential solutions for some of these challenges:

Pests and diseases: Integrated Pest Management (IPM) practices can be employed to reduce reliance on pesticides. IPM involves using a combination of techniques, such as biological control, crop rotation, and pest-resistant varieties, to manage pests and diseases effectively.

Labor shortages: Automation and agricultural robotics can help alleviate labor shortages by performing tasks such as harvesting and planting. Additionally, improving working conditions and providing better incentives for agricultural labor can attract more workers to the sector.

Economic challenges: Government support through subsidies and price stabilization measures can help farmers cope with economic challenges. Diversification of income sources and access to fair markets can also contribute to improving farmers' financial situations.

Data availability and quality: Initiatives to improve data collection and sharing in agriculture, such as the use of remote sensing and IoT devices, can enhance data availability. Efforts to ensure data quality through validation and verification processes are equally important.

Model interpretability: Explainable AI techniques can help in understanding how machine learning models make decisions. This can increase farmers' trust in the models and enable them to make more informed decisions based on the model's recommendations.

Awareness and education: Providing farmers with access to information, training, and resources can empower them to adopt sustainable and efficient agricultural practices. Government and non-government organizations can play a crucial role in conducting awareness campaigns and providing educational programs.

Addressing losses and waste in the food system: Implementing better post-harvest handling practices, investing in cold chain infrastructure, and improving supply chain logistics can help reduce losses and waste in the food system.

Crop damage by wild animals: Employing deterrent techniques such as fencing, noise makers, and scare devices can help mitigate crop damage by wild animals. In some cases, implementing conservation measures to protect natural habitats can also reduce human-wildlife conflicts.

It is essential to recognize that addressing these challenges requires collaboration among stakeholders, including governments, researchers, farmers, and the private sector. Technological innovations, coupled with supportive policies and community engagement, can contribute to building a more resilient and sustainable agricultural sector capable of meeting the growing demands of an increasing global population.

VII. Future Work/Ideas

These are excellent ideas for extending and enhancing the work on crop recommendation using machine learning. Each of these ideas addresses specific aspects of the agricultural domain and can provide valuable insights and benefits for farmers and stakeholders. Let's briefly discuss each idea:

1. Survey on Economic Impact: Conducting surveys with farmers to determine the cost savings and economic impact of using the crop recommendation models would provide valuable feedback on the practical benefits of the approach. This data can help in assessing the return on investment for adopting such technologies.

2. Mobile Application Implementation: Developing a user-friendly mobile application based on the crop recommendation models would bring these techniques directly to the end-users (farmers and agribusiness owners). A mobile app can offer real-time access to recommendations and other agricultural insights, making it more accessible and convenient for farmers.

3. Data Collection from Different Regions: Expanding the data collection to different regions and diverse agro-climatic conditions can help assess the generalizability and robustness of the crop recommendation models. This will enable the models to adapt to varying conditions and improve their accuracy.

4. Large Dataset Usage: Utilizing larger and more diverse datasets can improve the performance of machine learning models. A comprehensive dataset would capture a wide range of factors influencing crop yields, leading to more accurate and reliable recommendations.

5. Economic and Environmental Impact Assessment: Evaluating the economic and environmental impact of adopting the crop recommendation technique can provide a holistic view of its benefits. This analysis can demonstrate the cost-effectiveness and sustainability of using machine learning in agriculture.

6. Sensor-based Real-time Data Collection: Installing sensors on farms to collect real-time data on weather, soil conditions, and crop health can enhance the accuracy of the crop recommendation system. This approach enables farmers to make timely decisions and optimize resource usage.

By pursuing these extension ideas, researchers and developers can further refine and deploy the crop recommendation models, making them more practical and beneficial for farmers worldwide. Ultimately, the integration of machine learning technologies in agriculture can lead to improved productivity, sustainable practices, and better livelihoods for farmers.

VIII. Conclusion

In conclusion, this research paper has successfully demonstrated the effectiveness of crop recommendation models based on advanced machine learning algorithms and a deep neural network. The models developed in this study have the potential to revolutionize decision-making in the agricultural industry by providing valuable insights and recommendations for crop selection.

The positive implications of this research are multifaceted. Farmers can benefit from the accurate crop recommendations, enabling them to optimize their resources, improve yields, and make informed choices about their agricultural practices. Governments can leverage this technique to design policies and support programs that align with the needs of the agricultural sector, fostering sustainable growth and food security. Businesses in the agricultural domain can utilize the models to create innovative products and services that cater to the specific needs of farmers and contribute to the overall development of the industry. Additionally, by helping to stabilize agricultural goods prices, the technique can contribute to a more stable and resilient food supply chain.

Furthermore, the paper's detailed exploration of the challenges in agriculture sheds light on the real-world obstacles faced by farmers and stakeholders. Addressing these challenges through data-driven solutions, such as crop recommendation models, can significantly improve the agricultural landscape and contribute to the sector's growth and sustainability.

The presented future ideas for extension provide a roadmap for further research and development in this field. By exploring areas such as economic impact evaluation, real-time data collection through sensors, and data expansion across different regions, researchers can continue to refine and enhance crop recommendation models.

Overall, this research has made a substantial contribution to the agricultural domain. The scalable, accurate, and user-friendly nature of the proposed models makes them a valuable asset for various stakeholders in the agriculture sector. As technology continues to advance, these crop recommendation models have the potential to play a crucial role in shaping a more efficient, productive, and sustainable future for agriculture.

References

[1] A. L. Samuel, “Some studies in machine learning using the game of checkers,” IBM Journal of Research and Development, vol. 3, no. 3, pp. 210–229, 1959.

[2] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, “Machine learning basics,” Deep learning, pp. 98–164, 2016.

[3] G. Bonaccorso, Machine learning algorithms. Packt Publishing Ltd, 2017.

[4] A. Calzadilla, T. Zhu, K. Rehdanz, R. S. Tol, and C. Ringler, “Climate change and agriculture: Impacts and adaptation options in south africa,” Water Resources and Economics, vol. 5, pp. 24–48, 2014.

[5] T. Partap, “Hill agriculture: challenges and opportunities,” Indian Jour nal of Agricultural Economics, vol. 66, no. 902-2016-67891, 2011.

[6] T. O. Ayodele, “Types of machine learning algorithms,” New advances in machine learning, vol. 3, pp. 19–48, 2010.

[7] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, Applied logistic regression. John Wiley & Sons, 2013, vol. 398.

[8] R. E. Wright, “Logistic regression.” 1995.

[9] B. Charbuty and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, 2021.

[10] M. R. Segal, “Machine learning benchmarks and random forest regression,” 2004.

[11] K. Taunk, S. De, S. Verma, and A. Swetapadma, “A brief review of nearest neighbor algorithm for learning and classification,” in 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE, 2019, pp. 1255–1260.

[12] S. Chen, G. I. Webb, L. Liu, and X. Ma, “A novel selective naïve bayes algorithm,” Knowledge-Based Systems, vol. 192, p. 105361, 2020.

[13] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Sup port vector machines,” IEEE Intelligent Systems and their applications, vol. 13, no. 4, pp. 18–28, 1998.

[14] I. Steinwart and A. Christmann, Support vector machines. Springer Science & Business Media, 2008.

[15] K. Gurney, An introduction to neural networks. CRC press, 1997.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1–6.

[17] S. Sharma, S. Sharma, and A. Athaiya, “Activation functions in neural networks,” Towards Data Sci, vol. 6, no. 12, pp. 310–316, 2017.

[18] J. Dean and S. Ghemawat, “Mapreduce: Simplified data processing on large clusters,” in OSDI’04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, 2004, pp. 137–150.

[19] D. Dahiphale, R. Karve, A. V. Vasilakos, H. Liu, Z. Yu, A. Chhajer, J. Wang, and C. Wang, “An advanced mapreduce: Cloud mapreduce, enhancements and applications,” IEEE

Transactions on Network and Service Management, vol. 11, no. 1, pp. 101–115, 2014.

[20] R. Kiveris, S. Lattanzi, V. Mirrokni, V. Rastogi, and S. Vassilvitskii, “Connected components in mapreduce and beyond,” in SOCC 2014, 2014. [Online].

[21] D. Dahiphale, “Mapreduce for graphs processing: New big data algo rithm for 2-edge connected components and future ideas,” IEEE Access, vol. 11, pp. 54986–55001, 2023.

[22] A. Oikonomidis, C. Catal, and A. Kassahun, “Deep learning for crop yield prediction: a systematic literature review,” New Zealand Journal of Crop and Horticultural Science, vol. 51, no. 1, pp. 1–26, 2023.

[23] P. A, S. Chakraborty, A. Kumar, and O. R. Pooniwala, “Intelligent crop recommendation system using machine learning,” in 2021 5th Inter national Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 843–848.

[24] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, “Agroconsultant: Intel ligent crop recommendation system using machine learning algorithms,” in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–6.

[25] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, “Crop recommender system using machine learning approach,” in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1066–1071.

[26] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, “Crop recommendation system to maximize crop yield using machine learning technique,” International Research Journal of Engineering and Technology, vol. 4, no. 12, pp. 950–953, 2017.

[27] D. Reddy and M. R. Kumar, “Crop yield prediction using machine learning algorithm,” in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466–1470.

[28] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. Priya, “Prediction of crop yield using machine learning,” Int. Res. J. Eng. Technol.(IRJET), vol. 5, 2018.

[29] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, “Improving crop productivity through a crop recommendation system using ensembling technique,” in 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018, pp. 114–119.

[30] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, “Crop recommendation system for precision agriculture,” in 2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 32–36.

[31] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” Sensors, vol. 18, no. 8, p. 2674, 2018.

[32] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H. M. Ag goune, “Internet-of-things (iot)-based smart agriculture: Toward making the fields talk,” IEEE Access, vol. 7, pp. 129551–129583, 2019.

[33] V. Iglovikov, S. Mushinskiy, and V. Osin, “Satellite imagery feature de tection using deep convolutional neural network: A kaggle competition,” arXiv preprint arXiv:1706.06169, 2017.

[34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:1603.04467, 2016.

[35] S. Sharma, S. Sharma, and A. Athaiya, “Activation functions in neural networks,” Towards Data Sci, vol. 6, no. 12, pp. 310–316, 2017.

[36] B. Ding, H. Qian, and J. Zhou, “Activation functions and their character istics in deep neural networks,” in 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 1836–1841.

[37] J. Brownlee, “What is the difference between a batch and an epoch in a neural network,” Machine Learning Mastery, vol. 20, 2018.

[38] J. T. Barron, “A general and adaptive robust loss function,” in Pro ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4331–4339.

[39] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[40] A. Lydia and S. Francis, “Adagrad—an optimizer for stochastic gradient descent,” Int. J. Inf. Comput. Sci, vol. 6, no. 5, pp. 566–568, 2019.

[41] A. M. Taqi, A. Awad, F. Al-Azzo, and M. Milanova, “The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance,” in 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 140–145.

[42] P. Gulati, A. Sharma, and M. Gupta, “Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: A review,” Int. J. Comput. Appl, vol. 141, no. 14, pp. 19–25, 2016.

[43] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 233–240.

[44] T. B. Pathak, M. L. Maskey, J. A. Dahlberg, F. Kearns, K. M. Bali, and D. Zaccaria, "Climate change trends and impacts on california agriculture: a detailed review," Agronomy, vol. 8, no. 3, p. 25, 2018.

[45] N. Mancosu, R. L. Snyder, G. Kyriakakis, and D. Spano, "Water scarcity and future challenges for food production," Water, vol. 7, no. 3, pp. 975 992, 2015.

[46] E. Vallino, L. Ridolfi, and F. Laio, "Measuring economic water scarcity in agriculture: a cross-country empirical investigation," Environmental Science & Policy, vol. 114, pp. 73–85, 2020.

[47] A. Alam, "Soil degradation: a challenge to sustainable agriculture," International Journal of Scientific Research in Agricultural Sciences, vol. 1, no. 4, pp. 50–55, 2014.

[48] R. Lal, "Restoring soil quality to mitigate soil degradation," Sustainability, vol. 7, no. 5, pp. 5875–5895, 2015.

[49] M. Donatelli, R. D. Magarey, S. Bregaglio, L. Willocquet, J. P. Whish, and S. Savary, "Modelling the impacts of pests and diseases on agricultural systems," Agricultural systems, vol. 155, pp. 213–224, 2017.

[50] P. Alexander, C. Brown, A. Arneth, J. Finnigan, D. Moran, and M. D. Rounsevell, "Losses, inefficiencies and waste in the global food system," Agricultural Systems, vol. 153, pp. 190–200, 2017.

[51] B. Awasthi and N. B. Singh, "Status of human-wildlife conflict and assessment of crop damage by wild animals in gaurishankar conservation area, nepal," Journal of Institute of Science and Technology, vol. 20, no. 1, pp. 107–111, 2015.