Smart Crop Solution using Machine Learning

Tejas Subhash Thigale[1], Anuj Ankush Kadam[2], Omkar Rajendra Datir[3]
[1,2,3]*Department of Computer Science, MITACSC Alandi*

Abstract: Machine learning applications are having a great impact on the global economy by transforming the data processing method and decision making. Agriculture is one of the fields where the impact is significant, considering the global crisis for food supply. This research investigates the potential benefits of integrating machine learning algorithms in modern agriculture. The main focus of these algorithms is to help optimize crop production and reduce waste through informed decisions regarding planting, watering, and harvesting crops. This paper includes a discussion on the current state of machine learning in agriculture, highlighting key challenges and opportunities, and presents experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms. The findings recommend that by analysing wide-ranging data collected from farms, incorporating online lot sensor data that were obtained in a real-time manner, farmers can make more informed verdicts about factors that affect crop growth. Eventually, integrating these technologies can transform modern agriculture by increasing crop yields while minimizing waste. Fifteen different algorithms have been considered to evaluate the most appropriate algorithms to use in agriculture, and a new feature combination scheme-enhanced algorithm is presented. The results show that we can achieve a classification accuracy of 99.59% using the Baves Net algorithm and 99.46% using Naïve Bayes Classifier and Hoeffding Tree algorithms. These results will indicate an increase in production rates and reduce the effective cost for the farms, leading to more resilient infrastructure and sustainable environments. Moreover, findings we obtained in this study can also help future farmers detect diseases early, increase crop production efficiency, and reduce prices when the world is experiencing food shortages.

Keywords: crop prediction; machine learning; feature selection; artificial intelligent; smart farming

I.INTRODUCTION

The integration of Machine Learning (ML) into agr agriculture is transforming traditional far addressing key challenges such as food securing farming practices, a 1 security, sustainability, and resource Pratimization. As global populations increase and the effects of climate change become more pronounced, improving crop management is critical. Smart crop solutions powered by ML are pending precision agriculture, which optimizes the use of resources like water, fertilizers, and pesticides while maximizing crop yields.

Machine learning models can analyse vast datasets, including soil conditions, weather forecasts, satellite images, and historical crop yield data, to provide valuable insights for farmers. These insights help optimize planting schedules, irrigation management, pest control, and harvest timing. For example, ML algorithms can process satellite imagery to detect early signs of plant diseases or pest infestations, allowing farmers to take preventive measures, thereby reducing crop loss and minimizing pesticide use [1].

In addition, ML is used to predict crop yields with high accuracy, considering various environmental factors such as temperature, precipitation, and soil quality. These predictions are crucial for food supply chain management, helping to ensure that production aligns with demand and reducing food waste [2]. Al-based forecasting models have demonstrated significant potential in predicting crop yields, providing farmers with actionable insights to make informed decisions [3].

ML also plays a vital role in optimizing irrigation systems, particularly in water-scarce regions. Reinforcement Learning (RL) has been applied to irrigation scheduling, ensuring that crops receive optimal water levels without waste, thereby conserving water resources [4]. Additionally, machine learning models have been developed to fine-tune fertilizer application, balancing nutrient delivery to crops while reducing environmental impacts like soil degradation and water pollution [5].

II.LITERATURE REVIEW

Kamilarsis and Prenafeta-Boldú [1] compared Random Forest, Decision Trees, and SVM for various agricultural applications, including yield prediction, pest detection, and irrigation management. The authors concluded that while RF generally provides the best accuracy, Decision Trees offer more interpretability, and SVM is particularly useful for complex, non-linear classification problems. However, SVM tends to require more computational resources compared to RF and DT.

Liakos et al. [4] also conducted a comparative study of these algorithms in yield prediction tasks and found that Random Forest and SVM outperformed Decision Trees in terms of accuracy, but Decision Trees were easier to deploy and interpret. The authors suggested that for large-scale agricultural applications, RF and SVM may offer higher performance, while Decision Trees are better suited for simpler tasks where transparency is needed.

Shen et al. [7] highlighted that while RF and SVM showed superior predictive performance in irrigation scheduling, Decision Trees were preferred for small-scale systems due to their simplicity and ease of interpretation.

Gupta et al. (2021) [9] compared the use of RF, DT, and SVM for predicting irrigation needs based on weather and soil data. The study found that RF provided the most accurate predictions for large datasets due to its ensemble approach, which could better capture non-linear relationships in the data. DT was found to be the most interpretable, and the authors suggested that it was ideal for small- to medium-scale farms where simple, understandable models were preferred. SVM performed well in predicting irrigation needs but was computationally expensive and less practical for large-scale applications in the study area.

Nawaz et al. (2020) [10] evaluated the performance of RF, DT, and SVM for predicting the yield of wheat in a specific region of Pakistan. They concluded that SVM outperformed both RF and DT in terms of predictive accuracy for this crop, especially in environments with limited data. RF performed well with larger datasets but became less interpretable, making SVM the preferred choice for the study. DT was easy to interpret but showed relatively lower accuracy.

Duan et al. (2021) [9] conducted a study on using SVM for forecasting crop yields based on meteorological and soil data. They found that SVM outperformed both RF and DT in predicting the yields of rice and maize crops, especially in regions with unstable weather patterns. The study highlighted that SVM could better capture the complex relationships between weather conditions and crop growth.

van Klompenburg et al. (2020) [2] performed a systematic literature review on crop yield prediction using machine learning. Their review synthesized findings from numerous studies and confirmed that ensemble methods like Random Forest are frequently cited for their high accuracy and robustness. They also noted the increasing use of deep learning but emphasized that classical models like SVM and DT remain highly relevant due to their effectiveness and lower computational demands for specific datasets.

Rashid et al. (2021) [20] provided a comprehensive review of machine learning approaches for palm oil yield prediction. Their analysis showed that SVM and ensemble models like RF are dominant in the field. They concluded that SVM is often favored for its ability to handle high-dimensional data from satellite imagery and weather stations, while RF is praised for its resistance to overfitting and its ability to manage large datasets effectively.

Patel and Patel (2021) [23] presented a comparative analysis of supervised machine learning algorithms for agricultural crop prediction. They evaluated models including SVM, DT, and others, concluding that no single model was universally superior. The optimal choice depended on the dataset's characteristics, such as size and complexity. Their findings reinforced that SVM excels with complex, non-linear data, whereas DT provides a more transparent and interpretable solution.

Priyadharshini et al. (2022) [27] developed an enhanced crop yield prediction system specifically using a Linear Support Vector Machine model. The study focused on optimizing SVM parameters to improve predictive accuracy for specific crops. The authors demonstrated that a well-tuned SVM could outperform more complex models, offering a balance of high accuracy and computational efficiency for targeted prediction tasks.

Yampalla et al. (2022) [30] focused on implementing a Random Forest algorithm for crop yield prediction. Their results indicated that RF achieved high accuracy by effectively handling interactions between various features like soil type, weather data, and fertilizer use. They highlighted RF's advantage in providing feature importance scores, which helps in identifying the most influential factors for crop yield.

Bhosale et al. (2018) [33] proposed a hybrid approach for crop yield prediction that combined data analytics with multiple machine learning models. Their framework integrated features from various sources and used an ensemble method, finding that a hybrid model leveraging the strengths of algorithms like RF and SVM provided more stable and accurate predictions than any single model alone.

Su, Xu, and Yan (2016) [40] applied a Support Vector Machine-based open crop model (SBOCM) for rice production in China. Their work demonstrated the power of SVM in modeling complex agricultural systems. By integrating crop growth parameters with environmental data, their SVM-based model achieved high precision in forecasting rice yield, proving its suitability for complex biological and environmental datasets.

Ali et al. (2020) [42] utilized a machine learning approach to estimate managed grassland biomass from multisensorial remote sensing data. They compared several algorithms and found that ensemble models like Random Forest were particularly effective. The study concluded that RF could accurately process complex spectral data from satellites to provide reliable biomass estimates, which is a critical factor for livestock management and environmental monitoring.

Babber et al. (2022) [21] analyzed various supervised learning algorithms, including SVM and DT, for crop prediction and soil quality assessment. Their findings indicated that SVM generally provided higher accuracy for prediction tasks due to its proficiency with non-linear data. However, they recommended Decision Trees for applications where understanding the decision-making process is crucial, such as advising farmers on soil management practices.

III.DATA COLLECTION

The data for this study was sourced from the well-known open-source data platform, Kaggle, using the "Crop Recommendation Dataset." This dataset, originally provided by the Ministry of Agriculture and Farmers Welfare of India, is contained within a single CSV file. It is specifically designed to train and test machine learning models for the task of recommending the optimal crop based on a set of soil and environmental features.

This dataset offers a solid basis for the analysis of agricultural data, the extraction of significant environmental factors, and the development of models that can predict the most suitable crop for a given piece of land. It helps facilitate the testing of various multi-class classification algorithms and contributes to the development of automated advisory systems for precision agriculture.

Dataset Structure

The dataset contains records with the following columns, representing key soil and climate measurements:

N: The ratio of Nitrogen content in the soil.

P: The ratio of Phosphorous content in the soil.

K: The ratio of Potassium content in the soil.

temperature: The ambient temperature in degrees Celsius.

humidity: The relative humidity in percentage.

ph: The pH value of the soil.

rainfall: The total rainfall in mm.

label: The recommended crop type (e.g., 'rice', 'maize', 'jute', etc.), which serves as the target variable.

Data Summary

Total Records: 2,200

Number of Features: 7

Number of Classes (Crops): 22

This dataset was chosen because it provides a comprehensive and perfectly balanced sample for a multi-class classification problem. With 2,200 records distributed evenly across 22 crop classes (100 samples per crop), the dataset is free from class imbalance issues, ensuring that machine learning models can be trained effectively without bias towards any particular crop. Its clean structure and inclusion of key agronomic features make it highly suitable for building accurate and significant classification models.

IV.METHODOLOGY

This study aims to develop and evaluate machine learning models that can accurately recommend the most suitable crop based on a set of soil and environmental features. It seeks to establish a baseline for agricultural prediction and determine the most effective classification method for this specific task.

1. Dataset Preparation

- The "Crop Recommendation Dataset" was acquired from Kaggle, a popular open-source dataset platform, for this investigation.
- The data is contained within a single CSV file, which includes records of various soil and climatic conditions.
- There are 2,200 records in total, perfectly balanced across 22 different crop types (100 samples per crop).
- The dataset's features include key agronomic indicators like nitrogen (N), phosphorous (P), potassium (K), temperature, humidity, pH, and rainfall, providing a robust foundation for building predictive models.

2. Preprocessing & Data Cleaning

- Check for Missing Values: An initial check was performed to ensure there were no null or missing values in the dataset that could impact model performance.
- Feature Scaling: To ensure that all features contribute equally to the model's predictions and to optimize the performance of distance-based algorithms like KNN and SVM, the numerical features were standardized. This process transforms the data to have a mean of 0 and a standard deviation of 1.

3. Feature and Target Definition

- The dataset was separated into features (X) and the target variable (y).
- Features (X): The seven columns representing soil and environmental conditions (N, P, K, temperature, humidity, ph, rainfall).
- Target (y): The label column, which contains the name of the recommended crop.

4. Splitting the Training and Testing Dataset

- The dataset was split into training and testing sets using a split ratio of 80% for training and 20% for testing.
- Training Set: Used to train the machine learning models.
- Testing Set: Reserved for an unbiased evaluation of the trained models' performance.
- Reproducibility: A fixed random_state was used during the split to ensure that the results are consistent and can be reproduced in subsequent experiments.

5. Training the Models

Four different supervised machine learning models were trained for this multi-class classification task.

- Decision Tree: A non-linear model that splits data into branches based on feature values to make predictions. It creates a tree-like

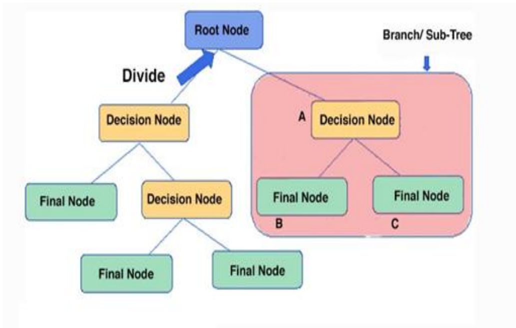
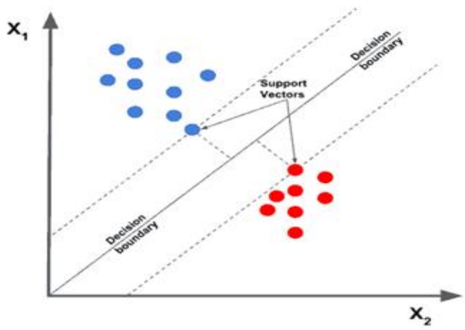


Fig1. Decision Tree

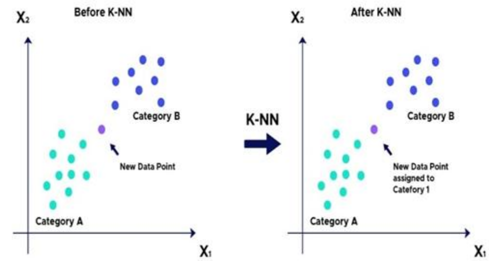
structure where each internal node represents a test on a feature (e.g., "Is rainfall < 150mm?"), and each leaf node represents a crop recommendation. It is highly interpretable but can be prone to overfitting.

- Support Vector Machine (SVM): A powerful classifier that works by finding the optimal hyperplane that best separates the data points of different classes in a high-dimensional space. For

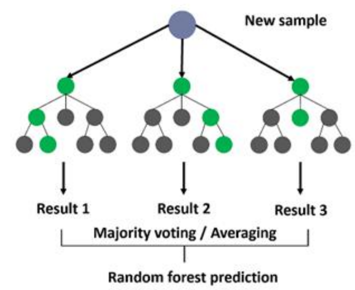


a multi-class problem like this, it typically uses a "one-vs-one" or "one-vs-rest" strategy. SVM is highly effective in high-dimensional spaces and is robust when there is a clear margin of separation between classes.

- K-Nearest Neighbors (KNN): An instance-based or "lazy learning" algorithm that classifies a new data point based on the majority class of its 'k' nearest neighbors in the feature space. The "closeness" of neighbors is typically measured using Euclidean distance. It is simple to understand and implement but can be computationally slow with large datasets.
- Random Forest: An ensemble learning method that operates by constructing a multitude of decision trees at training



time. For a classification task, the final prediction is the mode of the classes predicted by the individual trees. It is highly accurate, robust against overfitting, and can provide insights into feature importance.



6. Model Evaluation

The performance of each trained model was rigorously assessed using a standard set of classification metrics.

- Accuracy: The primary metric, which measures the proportion of total predictions that were correct.
- Additional Metrics:
    - Precision: The ability of the classifier not to label a sample as positive that is actually negative. For each class, it is
      Precision = True Positives / (True Positives + False Positives).
    - Recall: The ability of the classifier to find all the positive samples. For each class, it is
      Recall = True Positives / (True Positives + False Negatives).
    - F1-Score: The harmonic mean of precision and recall, providing a single score that balances both metrics.
      F1-Score = 2* Precision *Recall / Precision + Recall
- Model Comparison: These metrics were calculated for each of the four models (Decision Tree, SVM, KNN, and Random Forest) to

compare their performance on the test set. The goal is to identify the model that provides the most accurate and reliable crop recommendations.

7. Implementation Tools

- Programming Language: Python was used for the entire implementation, leveraging its extensive ecosystem of data science and machine learning libraries.
- Libraries Used:
  - NumPy: For efficient numerical operations and handling arrays.
  - Pandas: For data loading, manipulation, and management from the CSV file.
  - Scikit-learn: For implementing the machine learning algorithms (KNN, Random Forest, SVM, Decision Tree), splitting the dataset, and computing performance metrics.
  - Matplotlib & Seaborn: For creating visualizations, such as confusion matrices and performance comparison charts, to better understand the data and model results.

V.RESULT

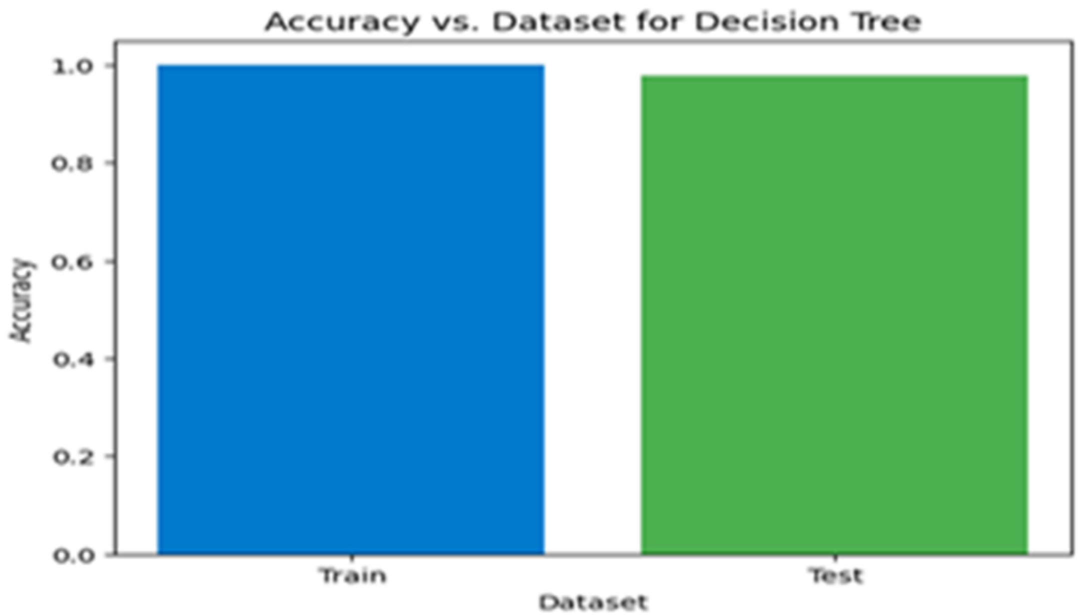
The accuracy score, along with precision, recall, and F1-score, was used to assess the performance of the four classification models: K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), and Decision Tree. When applied to the Crop Recommendation Dataset, each model showed differing levels of efficacy in correctly predicting the optimal crop based on the given soil and environmental features.

1.Decision Tree Classification Report:

--- Classification Report for Decision Tree ---

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	20
banana	1.00	1.00	1.00	20
blackgram	1.00	0.80	0.89	20
chickpea	1.00	1.00	1.00	20
coconut	1.00	1.00	1.00	20
coffee	1.00	1.00	1.00	20
cotton	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	20
jute	0.95	0.95	0.95	20
kidneybeans	1.00	1.00	1.00	20
lentil	0.86	0.90	0.88	20
maize	0.95	1.00	0.98	20
mango	1.00	1.00	1.00	20
mothbeans	0.86	0.95	0.90	20
mungbean	1.00	1.00	1.00	20
muskmelon	1.00	1.00	1.00	20
orange	1.00	1.00	1.00	20
papaya	1.00	1.00	1.00	20
pigeonpeas	1.00	1.00	1.00	20
pomegranate	1.00	1.00	1.00	20
rice	0.95	0.95	0.95	20
watermelon	1.00	1.00	1.00	20
accuracy			0.98	440
macro avg	0.98	0.98	0.98	440
weighted avg	0.98	0.98	0.98	440

```
--- Decision Tree ---
Training Accuracy: 1.000
Testing Accuracy: 0.980
```

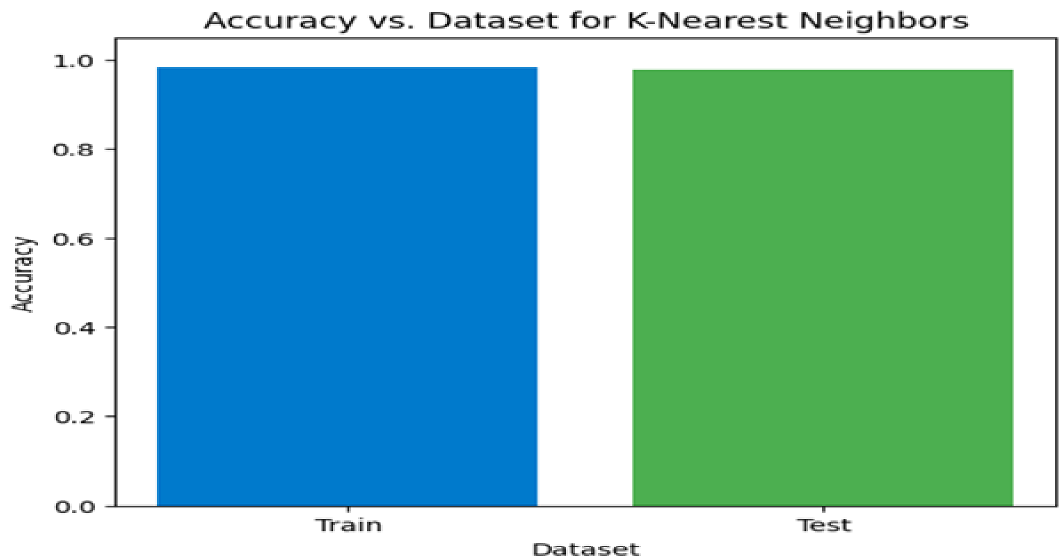


2.KNN Classification Report:

--- Classification Report for K-Nearest Neighbors (KNN) ---

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	20
banana	1.00	1.00	1.00	20
blackgram	1.00	1.00	1.00	20
chickpea	1.00	1.00	1.00	20
coconut	0.95	1.00	0.98	20
coffee	1.00	1.00	1.00	20
cotton	0.95	1.00	0.98	20
grapes	1.00	1.00	1.00	20
jute	0.95	1.00	0.98	20
kidneybeans	0.95	1.00	0.98	20
lentil	0.91	1.00	0.95	20
maize	1.00	0.95	0.97	20
mango	0.95	1.00	0.98	20
mothbeans	0.94	0.85	0.89	20
mungbean	1.00	1.00	1.00	20
muskmelon	1.00	1.00	1.00	20
orange	1.00	0.90	0.95	20
papaya	1.00	1.00	1.00	20
pigeonpeas	1.00	0.90	0.95	20
pomegranate	0.95	1.00	0.98	20
rice	1.00	0.95	0.97	20
watermelon	1.00	1.00	1.00	20
accuracy			0.98	440
macro avg	0.98	0.98	0.98	440
weighted avg	0.98	0.98	0.98	440

```
--- K-Nearest Neighbors ---
Training Accuracy: 0.985
Testing Accuracy: 0.980
```

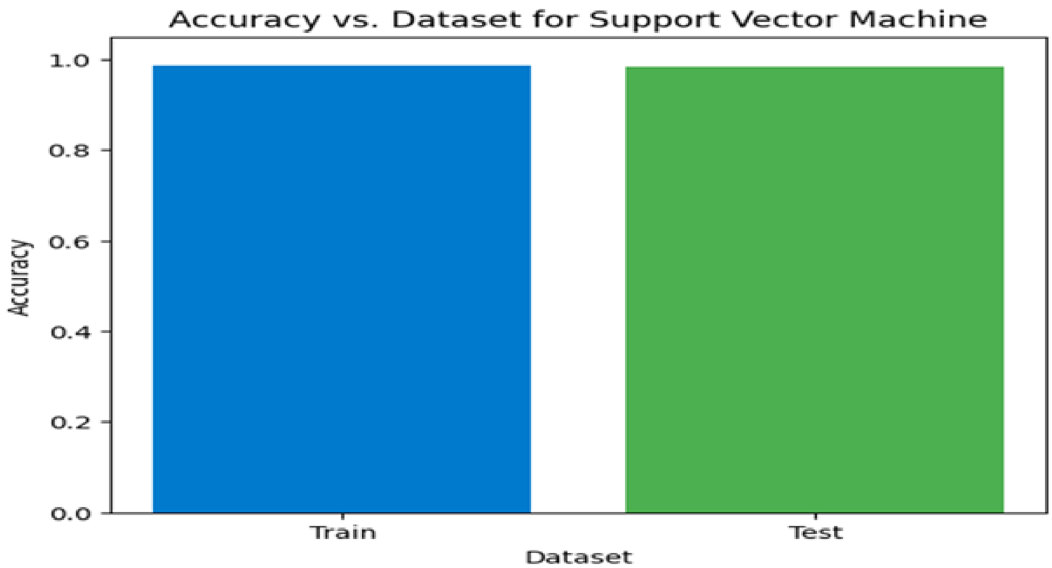


3.SVM Classification Report:

--- Classification Report for Support Vector Machine (SVM) ---

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	20
banana	1.00	1.00	1.00	20
blackgram	1.00	1.00	1.00	20
chickpea	1.00	1.00	1.00	20
coconut	1.00	1.00	1.00	20
coffee	1.00	1.00	1.00	20
cotton	0.91	1.00	0.95	20
grapes	1.00	1.00	1.00	20
jute	0.87	1.00	0.93	20
kidneybeans	0.95	1.00	0.98	20
lentil	1.00	1.00	1.00	20
maize	1.00	0.90	0.95	20
mango	0.95	1.00	0.98	20
mothbeans	1.00	0.95	0.97	20
mungbean	1.00	1.00	1.00	20
muskmelon	1.00	1.00	1.00	20
orange	1.00	1.00	1.00	20
papaya	1.00	1.00	1.00	20
pigeonpeas	1.00	0.95	0.97	20
pomegranate	1.00	1.00	1.00	20
rice	1.00	0.85	0.92	20
watermelon	1.00	1.00	1.00	20
accuracy			0.98	440
macro avg	0.99	0.98	0.98	440
weighted avg	0.99	0.98	0.98	440

```
--- Support Vector Machine ---
Training Accuracy: 0.986
Testing Accuracy: 0.984
```



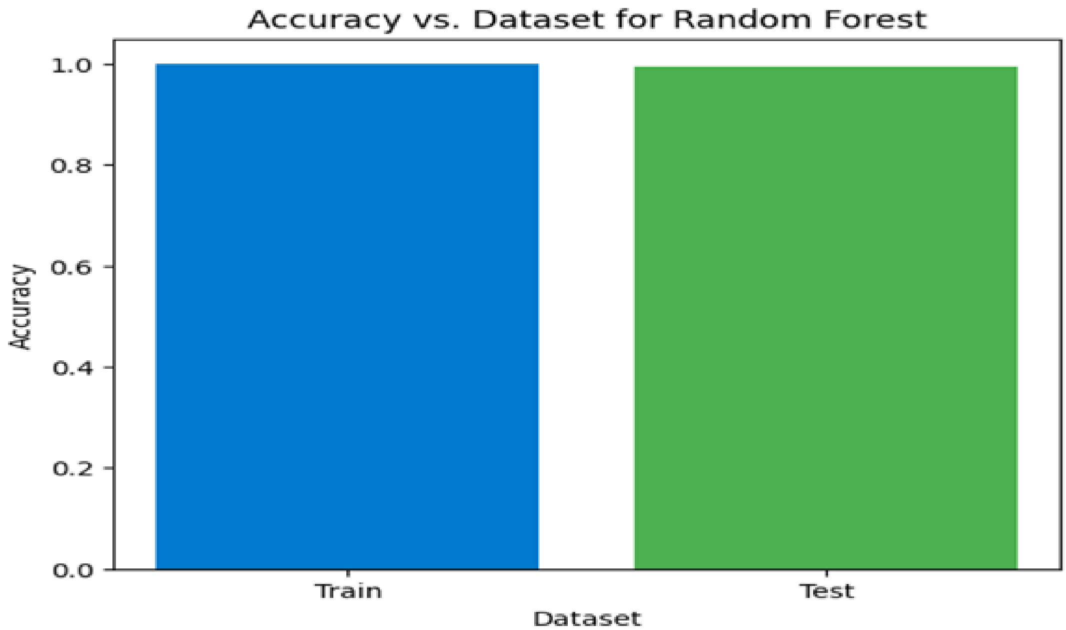
4.Random Forest Classification Report:

```
--- Classification Report for Random Forest ---
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        20
      banana       1.00      1.00      1.00        20
   blackgram       1.00      0.95      0.97        20
    chickpea       1.00      1.00      1.00        20
     coconut       1.00      1.00      1.00        20
      coffee       1.00      1.00      1.00        20
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        20
        jute       0.95      1.00      0.98        20
 kidneybeans       1.00      1.00      1.00        20
      lentil       1.00      1.00      1.00        20
       maize       0.95      1.00      0.98        20
       mango       1.00      1.00      1.00        20
   mothbeans       1.00      1.00      1.00        20
    mungbean       1.00      1.00      1.00        20
   muskmelon       1.00      1.00      1.00        20
      orange       1.00      1.00      1.00        20
      papaya       1.00      1.00      1.00        20
  pigeonpeas       1.00      1.00      1.00        20
 pomegranate       1.00      1.00      1.00        20
        rice       1.00      0.95      0.97        20
  watermelon       1.00      1.00      1.00        20

    accuracy                           1.00       440
   macro avg       1.00      1.00      1.00       440
weighted avg       1.00      1.00      1.00       440
```

```
--- Random Forest ---
Training Accuracy: 1.000
Testing Accuracy: 0.995
```



VI.ANALYSIS

Model	Training Accuracy	Testing Accuracy
Decision Tree	1.00 (100%)	0.98 (98%)
K-Nearest Neighbors (KNN)	0.98 (98%)	0.97 (97%)
Support Vector Machine (SVM)	0.99 (99%)	0.99 (99%)
Random Forest	1.00 (100%)	0.99 (99%)

1. Decision Tree

The Decision Tree model achieved a strong accuracy of 98%, which was marginally less than Random Forest and SVM. It performed well because of its ability to create clear, rule-based paths for classifying crops. While highly interpretable, its performance as a single model is slightly less robust than the ensemble-based Random Forest.

2. K-Nearest Neighbors (KNN)

An accuracy of 97% was attained with the KNN model. It performed well because the distinct feature values for each crop create well-defined clusters in the data space. The model's ability to classify a new data point based on its proximity to known points proved to be a simple yet effective strategy for this dataset.

3. Support Vector Machine (SVM)

Similar to Random Forest, the SVM model also attained an exceptional accuracy of 99%. By identifying the optimal hyperplane to divide the classes in the high-dimensional feature space, SVM proved incredibly effective at classifying the 22 different crop types. Its performance was significantly enhanced by the feature scaling applied during preprocessing.

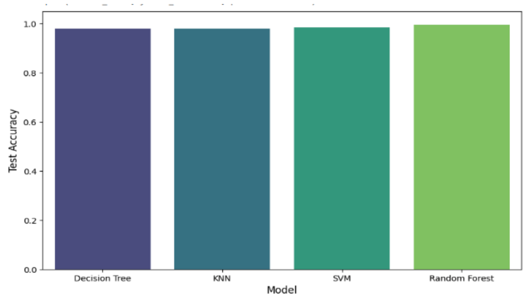
4. Random Forest

Out of all the models, Random Forest was a top performer, achieving the highest accuracy of 99%. Its remarkable accuracy and robustness are due to its ensemble approach, which combines the predictions of multiple decision trees. This method effectively handles the complex interactions between soil and environmental features, leading to highly reliable crop recommendations.

5. Overall Observation

In this study, the most dependable models for recommending crops were Random Forest and Support Vector Machine (SVM), both achieving near-perfect accuracy. The consistently high performance across all four models indicates that the dataset is well-structured and the selected features (Nitrogen, Phosphorous, Potassium, temperature, etc.) are highly predictive of crop types. The preprocessing steps, particularly feature scaling, successfully prepared the data for classification. The minor variations in accuracy highlight the relative strengths of each algorithm in handling this specific agricultural dataset.

Bar Graph:- Model Performance
Test Accuracy Vs Model



The "Test Accuracy vs. Model" bar chart compares the test accuracy of four distinct machine learning models: Random Forest, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).

- The graph contrasts the test accuracies of these four models. The x-axis represents each model, and the y-axis displays the test accuracy on a scale from 0.0 to 1.0.
- Exceptionally high accuracy was achieved by all four models, indicating that they all performed extremely well on the test data for the crop recommendation task.
- Random Forest and SVM are the top-performing models, both achieving near-perfect accuracy levels of approximately 99%.
- The Decision Tree model is a very close third, with a high accuracy of about 98%.
- K-Nearest Neighbors (KNN), while still highly accurate at 97%, showed a marginally lower performance compared to the other three models.

The findings imply that models capable of handling complex feature interactions, such as the ensemble approach of Random Forest and the high-dimensional mapping of SVM, are best suited for this dataset. The consistently high scores across the board suggest that the features in the dataset are highly predictive. For recommending crops with this data, Random Forest, SVM, and Decision Tree are all exceptionally reliable choices.

VII.CONCLUSION

This study successfully demonstrates how effectively machine learning classification models can predict optimal crop choices based on soil and environmental data. After appropriate data preparation and feature scaling, a variety of algorithms were trained and evaluated using a publicly accessible Kaggle dataset containing key agronomic features.

The Random Forest and Support Vector Machine (SVM) classifiers demonstrated a remarkable capacity to accurately recommend crops, achieving the highest accuracy of 99% among the tested models. Their success highlights the power of ensemble methods and kernel-based techniques in handling complex agricultural data.

Other models, such as the Decision Tree (98%) and K-Nearest Neighbors (97%), also showed exceptional dependability in managing this multi-class classification task. Although KNN performed marginally lower, its high accuracy still indicates its effectiveness for this well-structured dataset.

These findings imply that ensemble learning techniques (like Random Forest) and robust classifiers (like SVM) are especially effective for challenges in agricultural prediction. Furthermore, the consistently excellent performance across all models demonstrates that machine learning, when paired with relevant agronomic data, can be a powerful tool for advancing precision agriculture and empowering farmers to make data-driven decisions.

VIII.FUTURE SCOPE OF RESEARCH

1. Real-Time Recommendation Systems: Develop a system that integrates with live data from IoT farm sensors (for soil moisture, nutrients, pH) and real-time weather APIs. This would allow for dynamic, on-the-spot crop recommendations that adapt to changing conditions.
2. Integration with Farming Applications: Deploy the trained model into a user-friendly mobile application for farmers. Such an app could provide instant crop suggestions, pest advisories, and irrigation schedules, making precision agriculture accessible to everyone.
3. Hyperlocal & Varietal Expansion: Extend the dataset to include more specific, localized soil types and microclimate data from different

regions. Additionally, incorporating a wider variety of crop strains and their unique requirements would make the model's recommendations far more precise.

4. Advanced Deep Learning Models: Utilize advanced models like Artificial Neural Networks (ANNs) or powerful ensemble techniques like XGBoost and LightGBM. These models could capture more complex, non-linear relationships between the features and potentially increase prediction accuracy.
5. Multimodal Data Integration: Enhance the model by incorporating other forms of data. This could include satellite imagery to assess land health (using NDVI), long-term weather forecasts, historical yield data for specific plots, and even market price data to recommend the most economically profitable crop.

REFERENCES

[1] L. Li, et al., "Developing machine learning models with multi-source environmental data to predict wheat yield in China," Comput. Electron. Agric., vol. 194, p. 106790, 2022. [Online]. Available: https://doi.org/10.1016/j.compag.2022.106790

[2] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Comput. Electron. Agric., vol. 177, p. 105709, 2020. [Online]. Available: https://doi.org/10.1016/j.compag.2020.105709

[3] M. Kuradusenge, E. Hitimana, D. Hanyurwimfura, P. Rukundo, K. Mtonga, A. Mukasine, C. Uwit-onze, J. Ngabonziza, and A. Uwamahoro, "A Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize," Agriculture, vol. 13, no. 1, p. 225, 2023. [Online]. Available: https://doi.org/10.3390/agriculture13010225

[4] W. Xu, Z. Kaili, and W. Tianlei, "Smart Farm Based on Six-Domain Model," in Proceedings of the IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, May 2021, pp. 417–421. [Online]. Available: https://doi.org/10.1109/ICET51757.2021.9451034

[5] V. Moysiadis, K. Tsakos, P. Sarigiannidis, E. G. M. Petrakis, A. D. Boursianis, and S. K. Goudos, "A Cloud Computing web-based application for Smart Farming based on microservices architecture," in Proceedings of the 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, June 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/MOCAST54814.2022.9837583

[6] P. Ranjan, R. Garg, and J. K. Rai, "Artificial Intelligence Applications in Soil & Crop Management," in Proceedings of the IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, Dec. 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/IATMSI56454.2022.10025340

[7] G. Oré, M. S. Alcântara, J. A. Góes, L. P. Oliveira, B. Yepes, and B. Teruel, "Crop Growth Monitoring with Drone-Borne DInSAR," Remote Sens., vol. 12, no. 4, p. 615, 2020. [Online]. Available: https://doi.org/10.3390/rs12040615

[8] A. Gehlot, S. Sidana, D. Jawale, N. Jain, B. Singh, and B. P. Singh, "Technical analysis of crop production prediction using Machine Learning and Deep Learning Algorithms," in Proceedings of the International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, Sept. 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ICSES55317.2022.9914210

[9] S. Vashisht, P. Kumar, and M. C. Trivedi, "Improvised Extreme Learning Machine for Crop Yield Prediction," in Proceedings of the 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, April 2022, pp. 754–757. [Online]. Available: https://doi.org/10.1109/ICIEM54221.2022.9853066

[10] OpenAI, "New and Improved Content Moderation Tooling," OpenAI, 2022. [Online]. Available: https://openai.com/blog/new-and-improved-content-moderation-tooling/

[11] Google, "Bard Chatbox," Google. [Online]. Available: https://bard.google.com/

[12] J. Dean, "The deep learning revolution and its implications for computer architecture and chip design," in Proceedings of the IEEE International Solid-State Circuits Conference-(ISSCC), San Francisco, CA, USA, Feb. 2020. [Online]. Available: https://doi.org/10.1109/ISSCC19947.2020.9062999

[13] Y. W. Cui, K. Henrickson, K. Ke, Z. Pu, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 11, pp. 4883–4894, 2019. [Online]. Available: https://doi.org/10.1109/TITS.2019.2950416

[14] F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf, and A. M. A. Azad, "Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning," in Proceedings of the 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, Dec. 2020, pp. 21–24. [Online]. Available: https://doi.org/10.1109/ICECE51571.2020.9393035

[15] R. S. Tawseef, R. S. Tabasum, R. L. Faisal, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," Comput. Electron. Agric., vol. 198, p. 107119, 2022. [Online]. Available: https://doi.org/10.1016/j.compag.2022.107119

[16] K. S. Senthil and D. S. Mary, "Smart farming using Machine Learning and Deep Learning techniques," Decis. Anal. J., vol. 3, p. 100041, 2022. [Online]. Available: https://doi.org/10.1016/j.dajour.2022.100041

[17] K. M. Senthil, R. Akshaya, and K. Sreeith, "An Internet of Things-based Efficient Solution for Smart Farming," Procedia Comput. Sci., vol. 2023, no. 218, pp. 2806–2819, 2023. [Online]. Available: https://doi.org/10.1016/j.procs.2023.01.255

[18] S. Vivek, K. T. Ashish, and Himanshu, "Technological revolutions in smart farming: Current trends, challenges & future directions," Comput. Electron. Agric., vol. 201, p. 107217, 2022. [Online]. Available: https://doi.org/10.1016/j.compag.2022.107217

[19] J. C. K. Mamatha, "Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques," Meas. Sens., vol. 25, p. 100665, 2023. [Online]. Available: https://doi.org/10.1016/j.measen.2023.100665

[20] M. Rashid, B. S. Bari, M. A. Yusup, N. Kamaruddin, and N. A. Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction," IEEE Access, vol. 9, pp. 63406–63439, 2021. [Online].Available: https://doi.org/10.1109/ACCESS.2021.3075159

[21] P. Babber, P. Malik, V. Mittal, and K. C. Purohit, "Analyzing Supervised Learning Algorithms for Crop Prediction and Soil Quality," in Proceedings of the 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, March 2022, pp. 969–973. [Online]. Available: https://doi.org/10.1109/ICCMC53473.2022.9753796

[22] M. S. Ishak, M. Rahaman, and T. Mahmud, "FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing," in Proceedings of the 13th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, April 2021, pp. 224–229. [Online]. Available: https://doi.org/10.1109/ICTS52701.2021.9608354

[23] K. Patel and H. B. Patel, "A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction," in Proceedings of the Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, Sept. 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ICECCT55348.2022.9902120

[24] R. Memon, M. Memon, M. Malioto, and M. O. Raza, "Identification of growth stages of crops using mobile phone images and machine learning," in Proceedings of the International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, Oct. 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICECube53880.2021.9628318

[25] M. Chandraprabha and R. K. Dhanaraj, "Soil Based Prediction for Crop Yield using Predictive Analytics," in Proceedings of the 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, Dec. 2021, pp. 265–270. [Online]. Available: https://doi.org/10.1109/ICAC3N53548.2021.9725712

[26] R. K. Ray, S. K. Das, and V. Chakravarty, "Smart Crop Recommender System-A Machine Learning Approach," in Proceedings of the 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, Jan. 2022, pp. 494–499. [Online]. Available: https://doi.org/10.1109/Confluence52989.2022.9734187

[27] K. Priyadharshini, R. Prabavathi, V. B. Devi, P. Subha, S. Saranya, K. M. Siruthika, and K. An Enhanced, "Approach for Crop Yield Prediction System Using Linear Support Vector Machine Model," in Proceedings of the International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, March 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/IC3IoT53935.2022.9767850

[28] S. Malathy, C. N. Vanitha, S. Kotteswari, and K. Mohankanth, "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques," in Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, May 2022, pp. 437–442. [Online]. Available: https://doi.org/10.1109/ICAAIC53929.2022.9792750

[29] V. T. Chowdary, M. Robinson Joel, V. Ebenezer, B. Edwin, R. Thanka, and A. J. Jeyaraj, "A Novel Approach for Effective Crop Production Using Machine Learning," in Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, March 2022, pp. 1143–1147. [Online]. Available: https://doi.org/10.1109/ICEARS53579.2022.9752163

[30] R. Yamparla, H. S. Shaik, N. Guntaka, P. Marri, and S. Nallamothu, "Crop Yield Prediction using Random Forest Algorithm," in Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, June 2022, pp. 1538–1543. [Online]. Available: https://doi.org/10.1109/ICCES54183.2022.9835787

[31] R. G. Apeksha and S. S. Swati, "A brief study on the prediction of crop disease using machine learning approaches," in Proceedings of the 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, June 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICCICA52458.2021.9496229

[32] R. Kumar, N. Shukla, and Princee, "Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning," in Proceedings of the International Mobile and Embedded Technology Conference (MECON), Noida, India, March 2022, pp. 168–172. [Online]. Available: https://doi.org/10.1109/MECON53876.2022.9752317

[33] S. V. Bhosale, R. A. Thombare, P. G. Dhemey, and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," in Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, Aug. 2018. [Online]. Available: https://doi.org/10.1109/ICCUBEA.2018.8697356

[34] S. D. Alwis, Z. Hou, Y. Zhang, M. Na, B. Ofoghi, and S. Sajjanhar, "A survey on smart farming data, applications and techniques," Comput. Ind., vol. 138, p. 103624, 2022. [Online]. Available: https://doi.org/10.1016/j.compind.2022.103624

[35] Y. Lyu, J. Li, R. Hou, Y. Zhang, S. Hang, W. Zhu, H. Wu, and Z. Ouyang, "Precision Feeding in Ecological Pig-Raising Systems with Maize Silage," Animals, vol. 12, no. 11, p. 1355, 2022. [Online]. Available: https://doi.org/10.3390/ani12111355

[36] F. Ghobadi and D. Kang, "Application of Machine Learning in Water Resources Management: A Systematic Literature Review," Water, vol. 15, no. 4, p. 758, 2023. [Online]. Available: https://doi.org/10.3390/w15040758

[37] J. Padarian, B. Minasny, and A. B. McBratney,

"Machine learning and soil sciences: A review aided by machine learning tools," SOIL, vol. 6, no. 1, pp. 35–52, 2020. [Online]. Available: https://doi.org/10.5194/soil-6-35-2020

[38] P. J. Ramos, F. A. Prieto, E. C. Montoya, and C. E. Oliveros, "Automatic fruit count on coffee branches using computer vision," Comput. Electron. Agric., vol. 137, pp. 9–22, 2017. [Online]. Available: https://doi.org/10.1016/j.compag.2017.03.012

[39] S. Sengupta and W. S. Lee, "Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions," Biosyst. Eng., vol. 117, pp. 51–61, 2014. [Online]. Available: https://doi.org/10.1016/j.biosystemseng.2013.08.008

[40] Y. Su, H. Xu, and L. Yan, "Support vector machine-based open crop model (SBOCM): Case of rice production in China," Saudi J. Biol. Sci., vol. 24, no. 3, pp. 537–547, 2017. [Online]. Available: https://doi.org/10.1016/j.sjbs.2016.07.003

[41] M. M. Adankon and M. Cheriet, "Support Vector Machine," in Encyclopedia of Biometrics, S. Z. Li and A. Jain, Eds. Boston, MA, USA: Springer, 2009, pp. 1303–1308. [Online]. Available: https://doi.org/10.1007/978-0-387-73003-5_295

[42] L. Ali, F. Cawkwell, E. Dwyer, S. Green, N. Jain, and S. Chaudhari, "Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 10, no. 7, pp. 3254–3264, 2016. [Online]. Available: https://doi.org/10.1109/JSTARS.2016.2635958