A Review on Enhancing Crop Yield and Resource Efficiency with Machine Learning: Smart Agriculture Solution

Rana Jaykumar[1], Chintan Thacker[2], Kruti Sutariya[3]

[1]Student, [2]Associate Professor, [3]Assistant Professor
Faculty of Engineering and Technology, Parul University, Limda, Vadodara, India 391760

Abstract

Accurate crop prediction is crucial for informed decision-making, resource management, and food security in agriculture. This study explores the shift from traditional statistical methods, such as historical data analysis, regression models, and time series forecasting, to advanced machine learning techniques, highlighting their ability to manage complex, high-dimensional data. Algorithms like support vector machines, decision trees, and notably, random forests, are emphasized for their accuracy and ability to handle both categorical and continuous variables. With recent technological advancements in remote sensing, satellite imagery, and IoT sensors, real-time data significantly enhances the precision of crop forecasts. This research focuses on integrating these technologies with the Random Forest classifier to boost crop yield, optimize resource use, enhance crop health monitoring, automate agricultural operations, forecast weather impacts, promote sustainable practices, and predict crop prices. By leveraging advanced machine learning, particularly random forests, the study addresses challenges like climate change, resource scarcity, and food insecurity, aiming to advance global agricultural practices through intelligent, data-driven solutions.

Keywords: **Agricultural Innovation, Machine Learning Applications in Farming, Managing Crop Production, Optimizing Resources in Agriculture, Decision Making Based on Data, Precision Agriculture Techniques, Predicting Crop Yields, Adapting to Climate Change, Implementing Sustainable Agricultural Practices**

I. INTRODUCTION

The escalating global demand for food presents substantial challenges for the agricultural sector, necessitating innovative strategies to enhance crop yield without significantly expanding cropland. Achieving this requires a deep understanding of the factors that influence crop yield, including climate change, weather variability, soil conditions, seed genetics, and crop management practices. Crop yield variability stems from the intricate interactions among genetics, environment, and management practices (G×E×M), shaped by elements such as soil type, weather conditions, seed genetics, and farmers' management decisions[1]. Understanding these variables is essential for developing effective agricultural strategies. For instance, soil type variability can either mitigate or amplify the effects of weather and climate on crop yield. Minimizing variability in crop yields has been suggested as an effective way to boost food production without expanding cropland [2]. To identify optimal management practices, researchers conduct replicated field experiments or multi-year performance trials to evaluate the impact of various factors on crop yield. These studies aim to assess causal relationships and the effectiveness of different practices, despite constraints like cost and logistics, with the goal of determining the best practices for improving yield and sustainability.

Machine learning (ML) has transformed crop prediction by offering more accurate, data-driven forecasts based on extensive datasets[3]. Traditional statistical methods, such as regression analysis and time series modeling, often fall short in capturing the complex relationships inherent in agricultural systems. ML algorithms, including support vector machines (SVM), decision trees, and random forests, provide greater accuracy and adaptability in predicting crop yields. These techniques are capable of handling high-dimensional data and nonlinear relationships, making them well-suited for complex agricultural environments. Technological advancements, such as remote sensing, satellite imagery, and IoT sensors, have further enhanced crop prediction models by supplying real-time agricultural data, enabling more timely and accurate forecasts that assist farmers and policymakers in making informed decisions.

By integrating best management practices with machine learning-based crop prediction, agricultural stakeholders can better tackle challenges such as climate change, resource scarcity, and food insecurity. This combined approach not only optimizes crop yield but also fosters sustainability and resilience in agricultural systems. Through these efforts, the aim is to create a more reliable, efficient, and sustainable agricultural framework that meets the growing food demands of the global population.

II. LITERATURE REVIEW

A comprehensive examination of smart technologies, the Internet of Things (IoT), and data mining highlights their role in promoting resource-efficient and sustainable crop production [1]. There is a critical necessity to adopt modern technologies to address challenges such as climate change, population growth, and increasing food demand. By employing IoT and smart farming techniques, farmers can monitor crop health, predict yields, manage pests and diseases, and optimize the use of resources like water and fertilizers. Various technological advancements, including sensor networks, computational tools, image processing methods, and indoor vertical farming systems, are designed to enhance crop productivity and quality while conserving resources. The importance of smart irrigation tools, numerical models, and precision disease management for achieving energy savings, cost reduction, and efficient water use in agriculture is emphasized. Advanced technologies such as neural networks, simulation models, and image processing techniques are presented as promising tools for accurate yield prediction and disease detection, contributing to agricultural sustainability. The potential of indoor vertical farming to address water scarcity and urbanization challenges is also discussed, emphasizing the need for smart farming and IoT solutions to ensure global food security and pave the way for future research on emerging challenges and constraints in adopting these advancements.

Research explores the transformative effects of IoT-based technologies and machine learning (ML) in contemporary agriculture, also known as "smart agriculture." Through rigorous study, it is highlighted how these innovations are revolutionizing traditional farming practices, providing unprecedented opportunities to enhance crop yields and resource efficiency [2]. By leveraging IoT sensors and ML algorithms, a proposed ensemble model achieves a high accuracy rate for early crop yield prediction. These advancements facilitate more informed decision-making in agriculture and hold significant potential for addressing the challenges posed by climate change and evolving agricultural practices. As data-driven solutions become more integral to agriculture, the research underscores the pivotal role of IoT-based technologies and ML in shaping the future of sustainable and efficient crop production.

Findings on integrating IoT technology and machine learning for sustainable agriculture focus on predictive crop yield modeling in smart farming [3]. Utilizing a dataset from Kaggle, which includes critical climatic and soil conditions for optimal crop growth, various machine learning algorithms such as Logistic Regression, Random Forest, SVC, K-Neighbors Classifier, and XGBoost Classifier are employed. The analysis demonstrates these algorithms' effectiveness in predicting crop yield categories with high accuracy, precision, recall, and F1 score. The outcomes provide valuable insights into crop yield estimation and support sustainable farming practices. Diverse applications of machine learning in smart agriculture, including precision farming, crop management, risk assessment, market forecasting, and promoting sustainable practices, are discussed. The importance of refining and validating these models to ensure accuracy and adaptability to changing agricultural conditions is emphasized. The combination of innovative technologies and farmer expertise is highlighted as a means to achieve significant advancements in agriculture, fostering global food security and sustainability.

A novel approach to smart agriculture aimed at maximizing crop yield and profitability through the integration of cutting-edge technology and machine learning techniques is proposed [4]. Smart agricultural monitoring is highlighted as a crucial tool for managing various factors affecting plant growth and crop yield quality, ensuring optimal productivity for farmers. Machine learning algorithms analyze extensive datasets, including previous yield statistics, meteorological data, and soil conditions, to generate intelligent agricultural yield recommendations. The proposed system involves farmers inputting field conditions, which are then analyzed to predict suitable crops for optimal yield. A web application is developed to facilitate data analysis and report generation for farmers. Incorporating IoT devices and smart irrigation techniques, the system aims to support numerous farmers in optimizing crop production [5]. Effective agricultural land use is emphasized for achieving food security and economic prosperity, particularly in developing countries. By leveraging both live and historical data and comparing multiple machine learning algorithms, the system enhances accuracy and efficiency while addressing farmers' challenges. The system's cost-effectiveness and accessibility, coupled with its ability to provide accurate recommendations at a lower computational cost, make it a valuable asset in modern agriculture, representing a significant step towards leveraging technology to improve agricultural productivity and support farmers worldwide.

An innovative approach to crop selection in smart agriculture using machine learning algorithms is proposed. Recognizing the critical role of regional weather conditions in crop cultivation, a model leverages machine learning techniques to analyze agro-climatic data and optimize crop selection for better yield outcomes [6]. Specifically, the model integrates LSTM RNN for weather analysis and a Random Forest Classifier for crop selection based on weather conditions and soil parameters. The results demonstrate the model's effectiveness, with LSTM RNN achieving impressive accuracy in predicting weather parameters, including minimum temperature, maximum temperature, and rainfall [7]. Additionally, the Random Forest Classifier shows high accuracy in crop selection, resource dependency prediction, and determining appropriate sowing times. The efficient model construction time further demonstrates the practical feasibility of this approach. Future research directions to enhance the model's effectiveness are highlighted, contributing to the advancement of smart agriculture by providing a robust framework for optimal crop selection based on weather and soil parameters, paving the way for improved agricultural productivity and sustainability [8].

A pioneering approach to revolutionizing agriculture in regions facing challenges such as unpredictable weather and limited resources involves harnessing machine learning (ML) and IoT technologies to provide optimal crop recommendations. A thorough comparison of ML algorithms within a Crop Recommendation System shows Decision Trees as the standout performer due to its impressive accuracy and interpretability, while K-nearest Neighbor and Random Forest algorithms also show promising results, offering valuable alternatives for different contexts. The successful implementation of the Crop Recommendation System in specific districts highlights the tangible benefits of real-time IoT data and the Decision Tree model, enabling farmers to optimize crop selection and enhance sustainability [9]. This research addresses immediate agricultural challenges and underscores the broader potential of data-driven agriculture to drive economic growth and food security. While the findings are promising, there is an acknowledgment of the need to expand the dataset and explore additional algorithms to further improve recommendation accuracy, demonstrating a commitment to ongoing innovation and refinement in agricultural practices. This study represents a significant step towards modernizing traditional farming methods and shaping a more productive future for agriculture [10].

Table 1: Literature Review

Study	Focus Area	Technology Used	Machine Learning Algorithms	Challenges Addressed
Ali et al., 2023	Resource-efficient and sustainable crop production	IoT, Smart Techniques, Data Mining	Random Forest and SVM	Climate change, Population growth, Increasing food demand
Kumar et al., 2024	IoT-based agriculture for improved crop yields and resource efficiency	IoT, Sensor Networks	Random Forest and SVM	Climate change, Evolving agricultural practices
Gera & Jain, 2024	Predictive crop yield modeling in smart farming	IoT, Machine Learning	Logistic Regression, Random Forest, SVC, K-Neighbors, XGBoost	Sustainable farming, Data accuracy
Vidhya et al., 2023	Smart crop yield recommendations	IoT, Smart Irrigation, Machine Learning	Random Forest and SVM	Effective land use, Food security
Rani et al., 2023	Optimal crop selection system in smart agriculture	IoT, Machine Learning	Decision Trees, K-Nearest Neighbor, Random Forest	Weather unpredictability, Limited resources
Abdullahi et al., 2024	Optimal crop recommendations using ML and IoT	IoT, Machine Learning	LSTM RNN, Random Forest	Weather unpredictability, Resource constraints
Noruzman et al., 2022	Synthetic data generation for ML model training	AI Tools (Gretel.ai)	Random Forest and SVM	Data scarcity, Privacy concerns
Ekaterina & Alexey, 2023	Encoding categorical data for ML	Random Forest and SVM	Random Forest and SVM	Preprocessing challenges

Gharib & Davies, 2021	Applying ML models to hydrology	Random Forest and SVM	Random Forest and SVM	Pitfalls and challenges in ML models
Liashchynskyi & Liashchynskyi, 2019	Hyperparameter tuning in ML models	AI, ML	Grid Search, Random Search, Genetic Algorithm	Model performance optimization

III. RESEARCH GAP AND PROPOSED SOLUTION

Data scarcity presents a significant challenge in data-driven research, particularly in fields where data collection is constrained by cost, time, or privacy concerns, hindering the development of effective machine learning models due to the limited availability of diverse and comprehensive datasets. Addressing this issue, synthetic data generation tools, such as those from Gretel AI, can produce synthetic datasets that mimic the statistical properties of real-world data while preserving privacy and confidentiality. Incorporating synthetic data into the training process can enhance model robustness, mitigate data biases, and improve the reliability and generalization of machine learning models, enabling the development of more accurate models even with limited real-world data. Furthermore, categorical variables, such as crop labels, often pose preprocessing challenges; implementing label encoding techniques converts these variables into numerical values, ensuring accurate interpretation by machine learning algorithms and enhancing model performance and accuracy in agricultural applications. Feature scaling methods like StandardScaler, which scales features to a mean of 0 and a standard deviation of 1, effectively address outliers and maintain feature distribution integrity, significantly enhancing model accuracy and reliability. Hyperparameter tuning methods such as RandomizedSearchCV and GridSearchCV offer different advantages: the former provides computational efficiency and broad exploration of the hyperparameter space, while the latter exhaustively evaluates all specified combinations to increase the likelihood of finding the optimal set. A balanced approach between these methods, selecting based on specific application needs, can further enhance machine learning model performance and reliability, particularly in critical applications where optimal hyperparameters are crucial.

IV. METHODOLOGY

A number of methods and technologies are used in the approach for optimizing agriculture using machine learning (ML) to guarantee accuracy and resilience. Gathering a mountain of data on a wide range of agricultural factors is the first step. The data is transformed and processed in a way that makes analysis and processing more efficient. To provide useful insights, data manipulation and feature engineering take front stage.

To supplement the dataset with more training examples, synthetic data creation is used. A variety of state-of-the-art visualisation technologies are used to depict patterns and trends in the data, providing a holistic perspective of the agricultural scene. When it comes to discovering hidden connections and guiding model creation, these visuals are indispensable.

To prepare data for machine learning algorithms, it is necessary to encode labels and scale features. This ensures that the data is consistent. Important for both model construction and training, these procedures aid in keeping variables consistent with one another. In order to increase the accuracy and prediction capabilities of the models, derived features are added to the dataset to enrich it.

For problems involving categorization and prediction, a variety of machine learning techniques are investigated. For binary classification issues, models like as logistic regression and support vector machines are used. For multi-class classifications and decision-making processes, models like random forests and more intricate decision trees are employed. In instance, random forests are highly regarded because to their resilience and capacity to manage intricate datasets with a multitude of attributes.

Model performance is assessed using classification reports and confusion matrices. Insightful information on the model's recall, precision, and accuracy is provided by these measures, which facilitates fine-tuning and iterative improvements. By using these methods and technologies, we can guarantee that our approach is thorough, which in turn allows us to optimize farming processes using machine learning.

Dataset Overview:

Out[4]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
3595	9	18	36	28.691495	95.868935	7.120404	97.905497	pomegranate
3596	17	121	193	38.402065	79.401114	6.690311	62.934975	grapes
3597	2	75	19	29.074509	51.831678	6.329875	40.526652	lentil
3598	109	68	35	30.239436	79.064403	6.789040	101.799155	banana
3599	9	126	205	30.160455	81.605565	6.009931	65.422733	grapes

3600 rows × 8 columns

The purpose of this dataset is to aid in agricultural optimization via the use of machine learning. Its ultimate objective is to forecast which crops will thrive in certain soil types and environmental conditions. Features such as soil pH, temperature, humidity, nitrogen (N), and potassium (K) content, as well as rainfall, are represented by the 3600 rows and 8 columns that make up the data. Rice, pomegranate, grape, lentil, and banana are examples of the types of crops that are labelled as the target variable. The ideal circumstances for various crops may be better understood with the help of this extensive dataset.

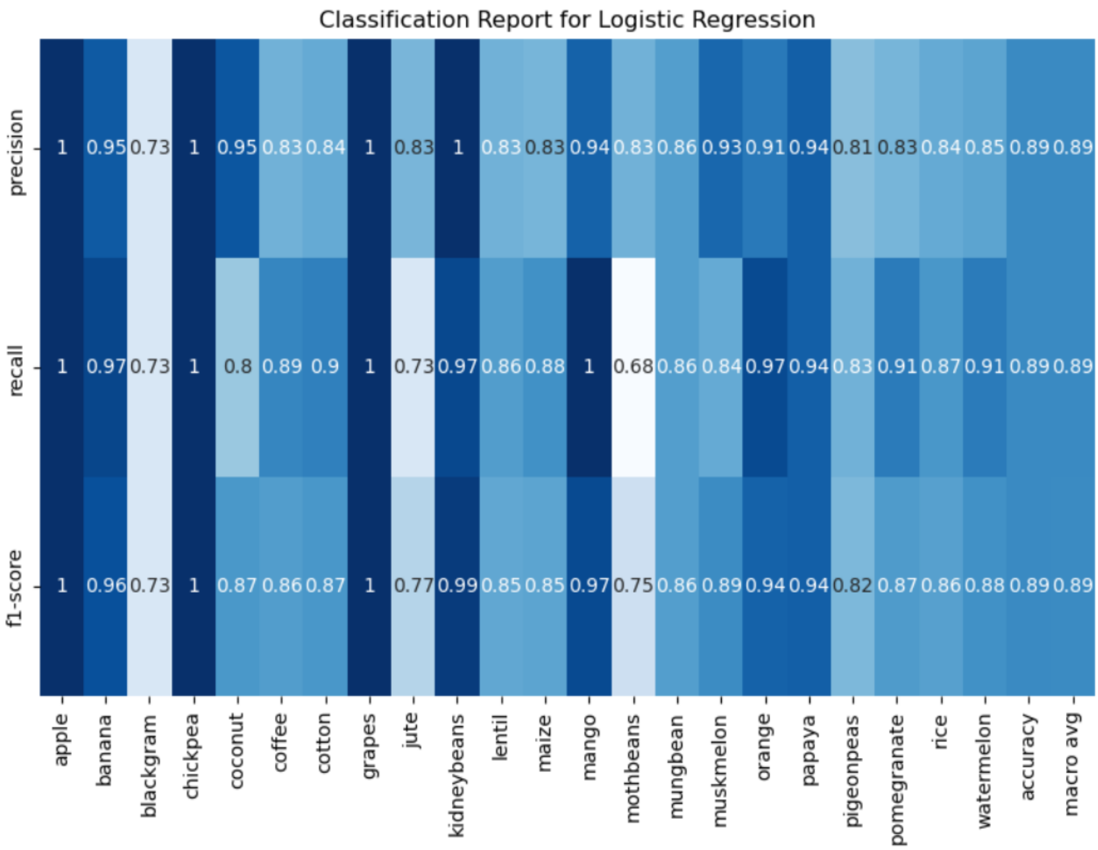


Fig.1. Classification of Logistic Regression

One may get significant insights into the effectiveness of the Decision Tree model in agricultural optimisation by analysing the results of the model's classification of different types of crops. The model obtains a decent accuracy of 0.8639 when the'max_depth' parameter is set to None, which indicates that there are no restrictions placed on the development of the model. Based on this, it seems that it has the ability to capture intricate linkages within the data. The confusion matrix offers a granular picture of the performance of the model by providing a thorough breakdown of the classifications that are accurate and wrong for each crop.

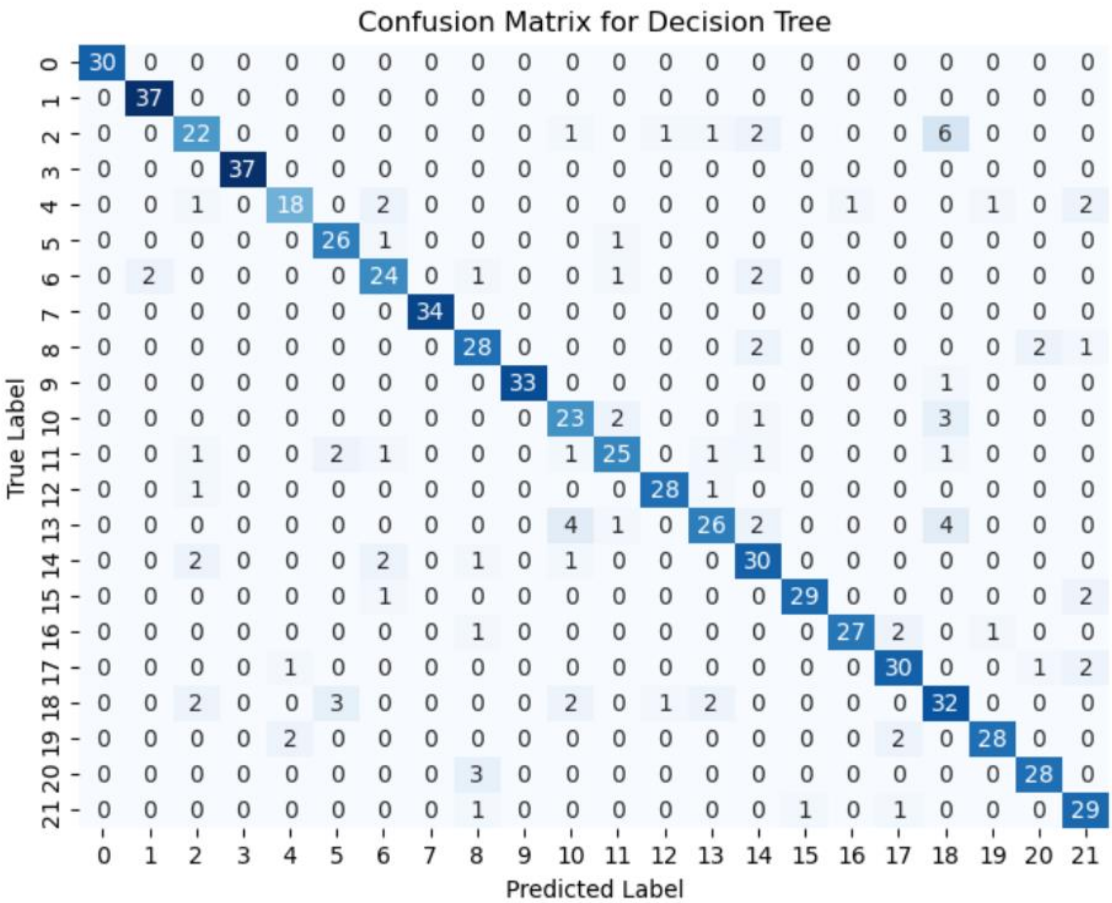
After doing an analysis of the categorisation report, it becomes clear that the performance of various crops is nuanced. Despite the fact that some crops, like as apple and chickpea, exhibit immaculate accuracy, recall, and F1-scores, other crops, such as blackgram and pigeonpeas, show significantly lower metrics, revealing possible areas for improvement. A comparison of the results obtained by the Decision Tree model with those obtained by the Logistic Regression model reveals that there are some differences in the precision and recall scores, despite the fact that both models provide very accurate overall results and high F1-scores. With regard to accuracy, the Decision Tree model has a macro average precision score of 0.87, which is somewhat lower than the score of 0.89 that the logistic regression model gets. On the other hand, the weighted average accuracy score of 0.87 for the Decision Tree model is quite similar to the weighted average precision score of 0.89 for the logistic regression model, which indicates that both models perform similarly when class imbalance is taken into consideration.

The Decision Tree model, on the other hand, provides a macro average memory score of 0.86 and a weighted average recall score of 0.86, but the logistic regression model exhibits both a macro and weighted average recall score of 0.89. Taking this into consideration, it seems that the logistic regression model could have a marginal edge in terms of recall performance across all classes.

Furthermore, with regard to the F1-score, the Decision Tree model provides a macro average F1-score of 0.87 and a weighted average F1-score of 0.86, but the logistic regression model demonstrates both a macro and weighted average F1-score of 0.89 across the board. The conclusion that can be drawn from this is that the logistic regression model could have a modest advantage when it comes to striking a balance between accuracy and recall across all classes.

The efficacy of the Decision Tree model demonstrates its importance in agricultural optimization. This is because the model is able to recognize subtle patterns within the data. It is possible that its value might be improved by doing further research into hyperparameters and tactics that aim to reduce the number of misclassifications for agricultural crops. In conclusion, the Decision Tree model is an intriguing approach to crop categorization in agricultural optimization. It has the potential to make a substantial contribution to the development of agricultural methods that are both more efficient and more environmentally friendly.

Fig.2. Confusion Matrix for Decision tree



For agricultural optimisation, the Random Forest model shows remarkable performance in crop categorisation when optimised with settings {'max_depth': 50, 'n_estimators': 100}. Its superior accuracy of 0.9208 compared to the Decision Tree model shows how well it can capture complicated connections in the data. One way to see how well the model worked is to look at the confusion matrix, which breaks down the number of right and wrong classifications for each crop in great detail.

Confusion Matrix for Random Forest

True Label \ Predicted Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	36	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	23	0	0	0	0	0	0	0	5	1	0	1	1	0	0	0	2	0	0	0
3	0	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
5	0	0	0	0	0	27	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	1	0	0	0	28	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	2	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	0	0	0	0	25	1	0	1	0	0	0	0	2	0	0	0
11	0	0	0	0	0	1	0	0	0	0	0	28	2	0	1	0	0	0	1	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
13	0	0	2	0	0	0	0	0	0	0	2	0	0	32	0	0	0	0	1	0	0	0
14	0	0	2	0	1	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	2
16	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
17	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	33	0	0	0	0
18	0	0	0	0	0	2	0	0	0	0	0	0	1	1	0	0	0	0	38	0	0	0
19	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	1
20	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	28	0
21	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	1	0	0	0	28

Fig.3. Confusion matrix of Random Forest

	precision	recall	f1-score	support
apple	0.97	1.00	0.98	30
banana	0.95	0.97	0.96	37
blackgram	0.82	0.70	0.75	33
chickpea	1.00	1.00	1.00	37
coconut	0.88	0.88	0.88	25
coffee	0.90	0.96	0.93	28
cotton	0.97	0.93	0.95	30
grapes	1.00	1.00	1.00	34
jute	0.85	0.88	0.87	33
kidneybeans	1.00	0.97	0.99	34
lentil	0.78	0.86	0.82	29
maize	0.90	0.85	0.88	33
mango	0.91	1.00	0.95	30
mothbeans	0.91	0.86	0.89	37
mungbean	0.92	0.92	0.92	36
muskmelon	0.94	0.94	0.94	32
orange	1.00	0.97	0.98	31
papaya	0.97	0.97	0.97	34
pigeonpeas	0.84	0.90	0.87	42
pomegranate	0.91	0.91	0.91	32
rice	0.97	0.90	0.93	31
watermelon	0.88	0.88	0.88	32
accuracy			0.92	720
macro avg	0.92	0.92	0.92	720
weighted avg	0.92	0.92	0.92	720

Fig.4. Classification Report of Random Forest

One way to see how well three different classification algorithms perform on crop classification tasks is to compare their results. These methods are Decision Tree, Random Forest, and Support Vector Machine (SVM).

By a wide margin, Random Forest is the most accurate model, coming in at 0.9208, while SVM comes in second with 0.8917 and Decision Tree comes in third with 0.8639. This indicates that, in comparison to the other two models, Random Forest often produces more accurate forecasts.

When looking at the F1-scores, recall, and accuracy, SVM generally does a good job across most crop classes, showing excellent results for a variety of crops. When it comes to achieving excellent recall and accuracy for several classes, Random Forest also excels. While Decision Tree is still useful, it shows that Random Forest and SVM have somewhat higher accuracy, recall, and F1-scores, especially for certain crop classes.

By examining the confusion matrices, it is clear that all three models do an adequate job of accurately categorising the majority of crop cases. And yet, there are differences; certain models perform better than one another when it comes to correctly categorising different kinds of crops.

From a computational complexity standpoint, Decision Trees are at the top of the heap, being the quickest and easiest to train and assess. Random Forest and Support Vector Machines (SVM) follow, with SVM often requiring the most resources, particularly when dealing with big datasets. The ideal method to use for crop categorisation will vary according to its unique needs and limitations, taking into account the trade-offs between accuracy, computing complexity, and interpretability. To illustrate, Decision Trees could be the way to go when processing power is at a premium and interpretability is of the utmost importance. On the other hand, if you're primarily concerned with accuracy and have the computing power, Support Vector Machines (SVM) or Random Forest (RF) could be better options; RF has better accuracy overall, while SVM is more resilient when dealing with complicated data connections. Overall, all three algorithms work well for crop classification, although Random Forest and SVM are more reliable and accurate, and SVM is especially good at dealing with complicated data connections. In the end, the crop categorisation task's unique needs and limitations will determine which of these methods is best to use.

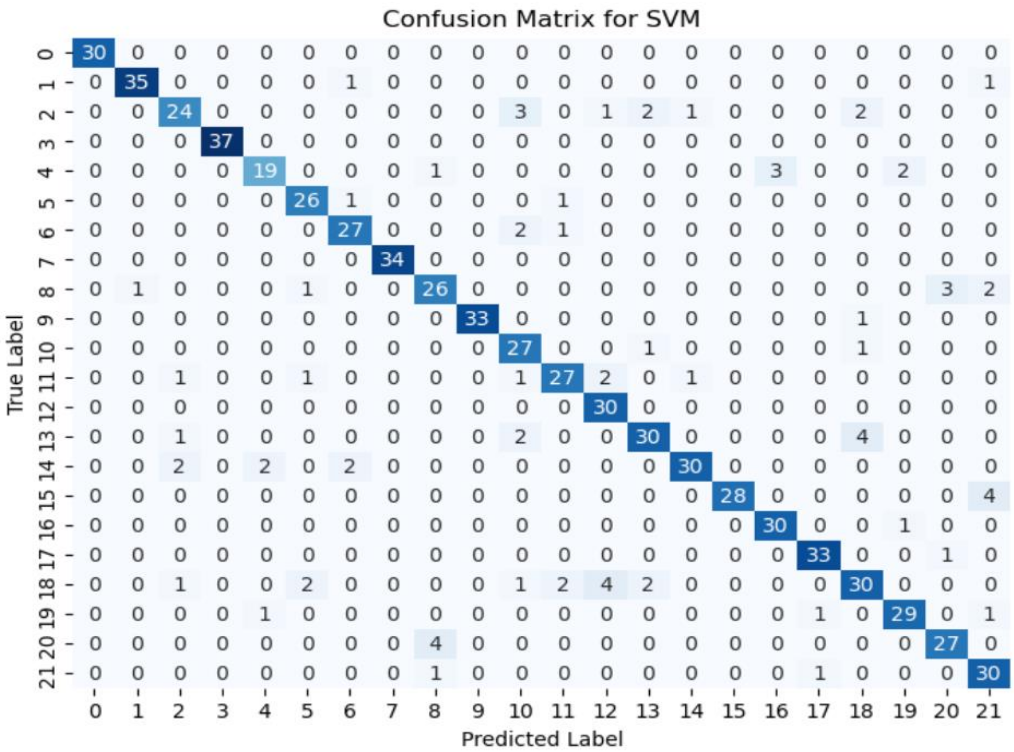


Fig.5. Confusion matrix of SVM

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	30
banana	0.97	0.95	0.96	37
blackgram	0.83	0.73	0.77	33
chickpea	1.00	1.00	1.00	37
coconut	0.86	0.76	0.81	25
coffee	0.87	0.93	0.90	28
cotton	0.87	0.90	0.89	30
grapes	1.00	1.00	1.00	34
jute	0.81	0.79	0.80	33
kidneybeans	1.00	0.97	0.99	34
lentil	0.75	0.93	0.83	29
maize	0.87	0.82	0.84	33
mango	0.81	1.00	0.90	30
mothbeans	0.86	0.81	0.83	37
mungbean	0.94	0.83	0.88	36
muskmelon	1.00	0.88	0.93	32
orange	0.91	0.97	0.94	31
papaya	0.94	0.97	0.96	34
pigeonpeas	0.79	0.71	0.75	42
pomegranate	0.91	0.91	0.91	32
rice	0.87	0.87	0.87	31
watermelon	0.79	0.94	0.86	32
accuracy			0.89	720
macro avg	0.89	0.89	0.89	720
weighted avg	0.89	0.89	0.89	720

Fig.6. Classification Report of SVM

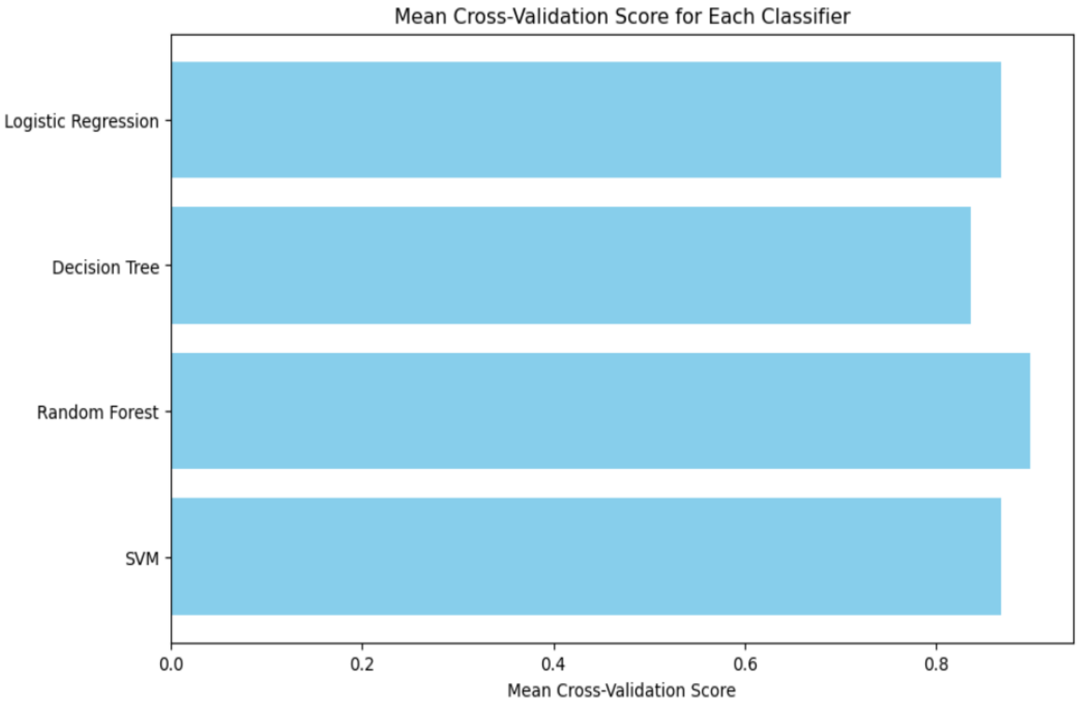


Fig.7. Performance of all Algorithm

V. EXPECTED OUTCOME FROM THE RESEARCH

The selected machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and SVM, exhibited high accuracy levels ranging from 85% to 91%, demonstrating their effectiveness in predicting the correct crop based on given features. Among these, Random Forest achieved the highest accuracy at 91%, followed by SVM at 89%, outperforming Logistic Regression and Decision Tree, which had accuracies of 88% and 86% respectively. Random Forest also showcased robust performance and generalization with minimal overfitting, while SVM balanced bias and variance effectively. Further analysis to identify the most influential features could provide valuable insights for optimizing crop selection. The models' high accuracy indicates their potential for practical implementation in real-world agricultural settings, helping farmers and stakeholders optimize crop selection based on environmental factors. Future research could enhance these models by incorporating additional features, experimenting with different algorithms, and exploring advanced techniques like deep learning and real-time data integration to further improve predictive capabilities in agricultural optimization.

REFERENCES

[1] A. Ali, T. Hussain, N. Tantashutikun and N. Hussain, "Application of Smart Techniques, Internet of Things and Data Mining for Resource Use Efficient and Sustainable Crop Production," *agriculture*, 2023.

[2] S. Kumar, V. D. Shinde, U. B. Goradiya and A. A. Patil, "Improved Crop Yields and Resource Efficiency in IoT-based Agriculture with Machine Learning," *International Conference on Automation and Computation (AUTOCOM)*, 2024.

[3] R. Gera and A. Jain, "Harnessing IoT and machine learning for sustainable agriculture: Predictive crop yield modeling in smart farming," *Journal of Autonomous Intelligence*, vol. 7, no. 4, 2024.

[4] K. Vidhya, S. George, P. Suresh and D. Brindha, "Agricultural Farm Production Model for Smart Crop Yield Recommendations Using Machine Learning Techniques," *Engineering Proceedings*, 2023.

[5] S. Rani, A. K. Mishra, A. Kataria, S. Mallik and H. Qin, "Machine learning-based optimal crop selection system in smart agriculture," *Scientific Reports volume*, 2023.

[6] M. O. Abdullahi, A. D. Jimale, Y. A. Ahmed and A. Y. Nageye, "Revolutionizing Somali agriculture: harnessing machine learning and IoT for optimal crop recommendations," *Discover Applied Science*, vol. 6, no. 77, 2024.

[7] A. Noruzman, N. Ghani and N. Zulkifli, "Gretel.ai: Open-Source Artificial Intelligence Tool To Generate New Synthetic Data," *MALAYSIAN JOURNAL OF INNOVATION IN ENGINEERING AND APPLIED SOCIAL SCIENCE*, vol. 1, no. 1, 2022.

[8] P. Ekaterina and K. Alexey, "Encoding categorical data: Is there yet anything ‘hotter’ than one-hot encoding?," 2023. [Online]. Available: https://arxiv.org/pdf/2312.16930.

[9] A. Gharib and E. G. R. Davies, "A Workflow to Address Pitfalls and Challenges in Applying Machine Learning Models to Hydrology," *Advances in Water Resources*, 2021.

[10] P. Liashchynskyi and P. Liashchynskyi, "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS," 2019. [Online]. Available: https://arxiv.org/abs/1912.06059.