ORIGINAL RESEARCH ARTICLE

Harnessing IoT and machine learning for sustainable agriculture: Predictive crop yield modeling in smart farming

Rashmi Gera[*], **Anupriya Jain**

Manav Rachna International Institute of Research and Studies, Faridabad 121004, India

[*] **Corresponding author:** Rashmi Gera, malik.reshu@gmail.com

ABSTRACT

The integration of Internet of Things (IoT) technology in agriculture has transformed traditional farming methods, enabling the emergence of smart agriculture systems. This research paper focuses on utilizing IoT and machine learning techniques to forecast crop yields based on climatic and soil conditions. The study utilizes a dataset sourced from Kaggle, which includes information on 22 distinct crops, such as Maize, Wheat, Mango, Watermelon, and others. The dataset encompasses crucial climatic factors like temperature, humidity, and rainfall, as well as essential soil conditions necessary for optimal crop growth. By employing advanced machine learning algorithms on this dataset, the objective is to develop accurate models capable of predicting crop yields. This, in turn, assists farmers in making well-informed decisions regarding crop management and optimizing agricultural productivity. The outcomes of this research have significant implications for the agricultural industry, providing valuable insights into crop yield estimation and supporting the implementation of sustainable farming practices.

Keywords: IoT; machine learning; sustainable agriculture; smart farming; predictive modeling; crop yield estimation

1. Introduction

Agriculture plays a pivotal role in sustaining human life by providing food, fiber, and raw materials for various industries. However, the challenges facing the agricultural sector are increasing due to a growing global population, changing climatic conditions, and limited availability of resources. In this context, the concept of sustainable agriculture has gained significant attention as a means to ensure long-term food security while minimizing environmental impact.

Sustainable agriculture encompasses practices that promote the efficient use of resources, preservation of ecosystems, and the adoption of socially responsible approaches. It aims to strike a balance between agricultural productivity, environmental conservation, and socioeconomic well-being. Achieving sustainable agriculture requires innovative solutions that optimize resource utilization, enhance productivity, and mitigate negative environmental consequences.

In recent years, advancements in technology, particularly in the areas of Internet of Things (IoT) and machine learning, have revolutionized various industries. These technologies have the potential to revolutionize agriculture as well, leading to the emergence of smart farming systems. Smart farming combines the power of IoT devices, sensors, data analytics, and machine learning

algorithms to enable precision agriculture, real-time monitoring, and data-driven decision-making.

The integration of IoT in agriculture has enabled the collection of vast amounts of data from various sources, including soil sensors, weather stations, and crop monitoring devices. This data provides valuable insights into environmental conditions, soil quality, crop health, and water usage. Machine learning algorithms can then be applied to analyze this data and develop predictive models for crop yield estimation, disease detection, and optimal resource allocation.

The main objective of this research paper is to explore the potential of IoT and machine learning technologies in achieving sustainable agriculture through predictive crop yield modeling in smart farming systems. By accurately predicting crop yields, farmers can optimize their resource allocation, make informed decisions, and enhance overall productivity. This, in turn, contributes to sustainable agricultural practices by minimizing resource waste, reducing environmental impacts, and ensuring economic viability.

The research will investigate various aspects of utilizing IoT and machine learning in smart farming for sustainable agriculture. It will delve into the methodologies employed, the challenges encountered, and the potential opportunities for improving crop yield prediction and resource optimization. Case studies and real-world implementations will be examined to showcase successful applications of these technologies in diverse agricultural contexts.

Furthermore, the research will address the challenges associated with the adoption of IoT and machine learning in agriculture, such as data quality, scalability, and farmer acceptance. Strategies for overcoming these challenges will be explored, along with considerations for policy implications and capacity building to enable widespread adoption of sustainable agricultural practices.

The findings of this research will have implications for farmers, policymakers, and stakeholders in the agriculture industry. By harnessing the potential of IoT and machine learning in smart farming, sustainable agricultural practices can be advanced, contributing to a more efficient and environmentally conscious food production system. The research aims to provide valuable insights, recommendations, and guidelines for stakeholders to effectively implement and leverage these technologies for sustainable agriculture.

In summary, this research paper seeks to explore the potential of IoT and machine learning in achieving sustainable agriculture through predictive crop yield modeling in smart farming systems. By harnessing these technologies, it is possible to optimize resource utilization, enhance decision-making, and foster long-term sustainability in agriculture. The subsequent sections of the paper will delve into the methodologies employed, present the results and analysis, discuss the implications, and provide recommendations for future research and implementation.

2. Literature review

The following literature review examines relevant studies and research papers contributing to the understanding and application of smart agriculture, crop yield prediction, and the use of IoT technology in agriculture.

These studies provide valuable insights into the application of IoT, machine learning, and data analysis techniques in smart agriculture, crop yield prediction, and agricultural resource management. By integrating these methodologies and approaches, researchers and farmers can harness the power of data-driven models to enhance agricultural productivity, optimize resource utilization, and promote sustainable farming practices.

Literature review in tabular form as shown in **Table 1**.

Table 1. Literature review.

Study	Key findings
Kanchan and Shardoor[1]	Introduced "Krashignyan", an IoT-based farmer support system for informed crop management. Leveraged IoT technology to collect and analyze environmental data for improved agricultural practices.
Javed et al.[2]	Conducted a comparative review of IoT operating systems, networking technologies, applications, and challenges. Explored the technical aspects of IoT systems and their potential applications in agriculture.
Issad et al.[3]	Provided a comprehensive review of data mining techniques in smart agriculture. Explored the application of data mining for crop yield prediction and crop disease detection.
Flak[4]	Presented an overview of technologies for sustainable biomass supply, relevant to smart agriculture. Discussed various technologies and their potential for sustainable agricultural practices.
Kavita[5]	Focused on satellite-based crop yield prediction using machine learning algorithms. Utilized satellite imagery and machine learning to estimate crop yields.
Mupangwa et al.[6]	Conducted a temporal rainfall trend analysis in southern Africa. Emphasized the importance of understanding rainfall patterns for optimized agricultural practices
Gado Djibo et al.[7]	Investigated statistical models for seasonal rainfall forecasts in a specific region. Demonstrated the potential of statistical models in aiding decision-making for crop selection and planning.

These studies collectively contribute to the understanding and application of smart agriculture, crop yield prediction, and the utilization of IoT technology in agriculture. Each study offers unique insights into various aspects, including IoT systems, data mining techniques, biomass supply, satellite-based prediction, rainfall trends, and statistical models. These findings are valuable for optimizing agricultural practices, enhancing productivity, and promoting sustainable agriculture.

3. Methodology used

The methodology employed in this research paper for crop yield prediction combines IoT data collection, data preprocessing, feature selection, and machine learning algorithms. The following steps outline the approach taken.

Data collection: The dataset used in this study is obtained from Kaggle and consists of information on 22 unique crops, including climatic and soil conditions necessary for optimal crop growth[8–12]. Parameters such as temperature, humidity, rainfall, and soil conditions are collected in real-time using IoT devices and sensors.

Data preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for analysis. Steps such as data normalization or scaling may be applied to bring the features to a similar scale, aiding in the accuracy of the subsequent machine learning models.

Feature selection: As the dataset may contain numerous features, not all of them may significantly contribute to crop yield prediction. Feature selection techniques are employed to identify the most relevant and informative features, reducing the dimensionality of the dataset and enhancing the efficiency and accuracy of the machine learning models.

Machine learning model selection: Crop yield prediction can be accomplished using a range of machine learning algorithms, with the specific choice dependent on the data characteristics and problem at hand. Commonly used algorithms include linear regression, decision trees, random forests, support vector machines, and neural networks[13,14]. The selection of the most suitable model is guided by evaluating their performance metrics, such as accuracy, precision, recall, and F1-score.

Model training and evaluation: The dataset is divided into a training set and a testing set. The training set is used to train the machine learning model by providing it with input features and their corresponding crop

yield values. This process allows the model to learn the underlying patterns and relationships. The trained model is then evaluated using the testing set to assess its performance and generalization ability to unseen data. Evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or the coefficient of determination (R-squared) are employed to measure the model's predictive accuracy.

Model optimization and fine-tuning: Based on the performance of the initial model, further optimization and fine-tuning may be conducted. This can involve hyperparameter tuning, regularization techniques, or ensemble methods to enhance the model's predictive capabilities and mitigate overfitting.

Crop yield prediction: The trained and optimized model utilizes the input features to generate predictions of crop yields, providing valuable insights for farmers to make informed decisions regarding crop planning, resource allocation, and risk management.

Through this methodology, the research aims to develop accurate crop yield prediction models by leveraging IoT data and machine learning algorithms. These models will enable farmers to optimize their agricultural practices, improve productivity, and make informed decisions for sustainable crop management. The research findings will contribute to the advancement of precision agriculture and sustainable farming practices. The distributions of parameters such as rainfall, pH, humidity, and temperature are visually presented in **Figures 1–7** to provide a better understanding of their variations and relationships.

The distribution analysis helps us understand how this particular agricultural condition is distributed or scattered across different regions, fields, or time periods. It allows us to identify patterns, trends, and variations in the condition, which can be valuable for farmers, researchers, and policymakers to make informed decisions and optimize agricultural practices.

For example, let's say "Agriculture Condition N" represents soil pH level, which is a critical factor influencing nutrient availability to plants. By studying the distribution of soil pH across various farmlands in a region, we can identify areas where the soil is too acidic, too alkaline, or within the optimal pH range for specific crops. This information can guide farmers in choosing suitable crops for each region or taking appropriate measures like soil amendments to improve crop productivity.

Distribution analysis often involves plotting the data on a graph or chart, such as a histogram or a frequency distribution curve, to visualize the spread and frequency of the specific condition across the dataset. From the resulting visual representation, we can draw insights regarding the central tendency, variability, outliers, and potential correlations with other factors.

Based on the provided information, the research aims to develop accurate crop yield prediction models using IoT data and machine learning algorithms. The researchers intend to leverage data from various sensors and devices connected through the Internet of Things (IoT) to collect real-time information about agriculture conditions such as rainfall, pH, humidity, and temperature. By utilizing advanced machine learning algorithms, they aim to analyze these data sets and develop models that can predict crop yields accurately.

Figures 1–7 in the research likely represent visual presentations of the distributions of the mentioned agricultural conditions (rainfall, pH, humidity, and temperature). Each figure may display the variation and relationship of a specific parameter across different regions, time periods, or agricultural practices.

Figure 1 agriculture condition N: This figure represents the distribution of "Agriculture Condition N" across the study area or dataset. As mentioned earlier, "Condition N" could be any specific parameter that significantly impacts crop growth and productivity. By visually presenting the distribution of "Condition N", researchers can identify patterns and variations, allowing farmers to understand the prevalence of this condition in different regions and make informed decisions for optimizing agricultural practices.

Figure 2 agriculture condition P: Similarly, this figure illustrates the distribution of "Agriculture Condition P" across the dataset. "Condition P" could be another critical parameter, such as soil phosphorus content, which is vital for root development and flowering in plants. Analyzing its distribution can help farmers assess the phosphorus availability in different fields and guide them in applying appropriate fertilizers or soil amendments.

Figure 3 agriculture condition K: This figure shows the distribution of "Agriculture Condition K", which could represent soil potassium content. Potassium is essential for enzyme activation and overall plant health. Understanding its distribution can assist farmers in maintaining proper potassium levels in the soil and enhancing crop growth and productivity.

Figure 4 distribution for rainfall conditions: This figure visually presents the distribution of rainfall conditions across the study area or over a specific period. It provides insights into the frequency and intensity of rainfall, which is crucial for proper irrigation planning and water management. Farmers can use this data to adjust irrigation schedules and adapt their farming practices based on rainfall patterns.

Figure 5 distribution for pH conditions: This figure displays the distribution of pH levels in the soil. As previously mentioned, pH plays a vital role in nutrient availability to plants. Analyzing its distribution helps farmers identify areas with soil that is too acidic or alkaline, enabling them to take corrective measures and optimize pH levels for different crops.

Figure 6 distribution for humidity conditions: This figure showcases the distribution of humidity levels in the atmosphere. Humidity affects plant transpiration and water stress, making it essential for farmers to understand its variations. By studying this distribution, farmers can make informed decisions about irrigation, particularly during periods of low humidity when plants are prone to water stress.

Figure 7 Distribution for temperature conditions: This figure presents the distribution of temperature across the study area or during different growing seasons. Temperature has a significant impact on crop growth and development. The figure highlights areas or periods with high and low temperatures, aiding farmers in choosing suitable crop varieties and planning planting times.

These visual representations (**Figures 1–7**) collectively contribute to the research's goal of developing accurate crop yield prediction models by leveraging IoT data and machine learning algorithms. The findings obtained from analyzing these distributions will help farmers optimize their agricultural practices, improve productivity, and make informed decisions for sustainable crop management, advancing the field of precision agriculture.
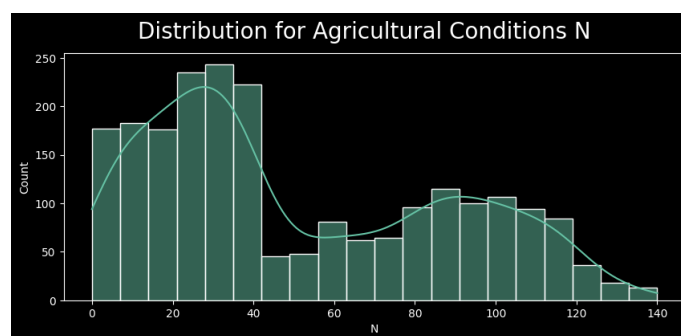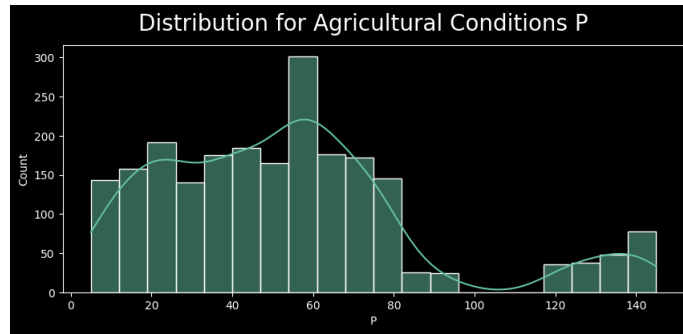

Figure 1. Agriculture condition N.
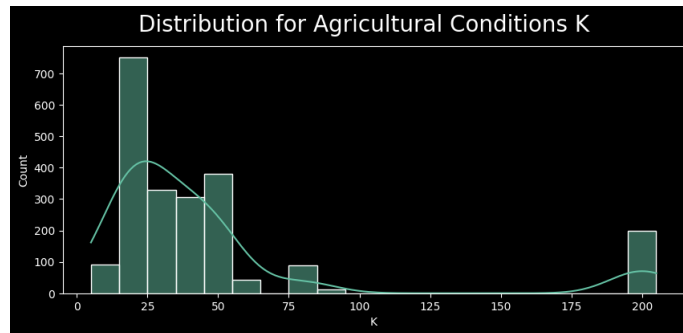
Figure 2. Agriculture condition P.

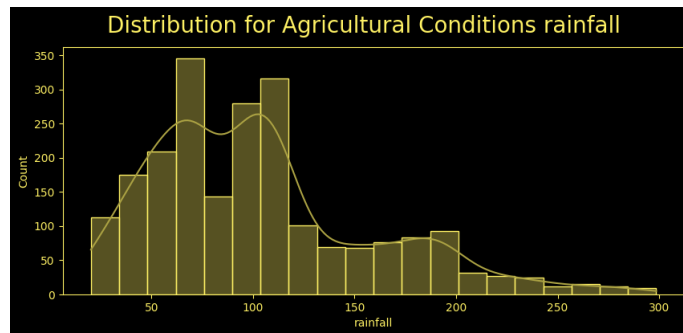
Figure 3. Agriculture condition K.

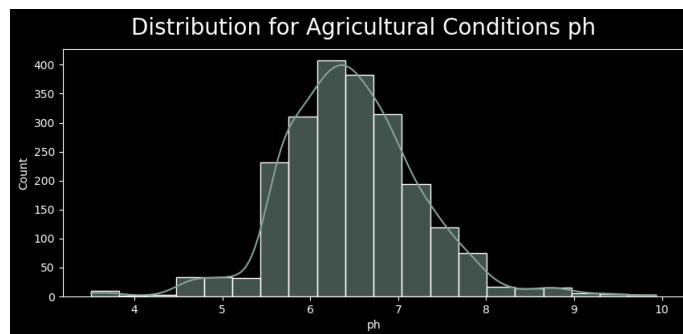**Figure 4.** Distribution for rainfall conditions.

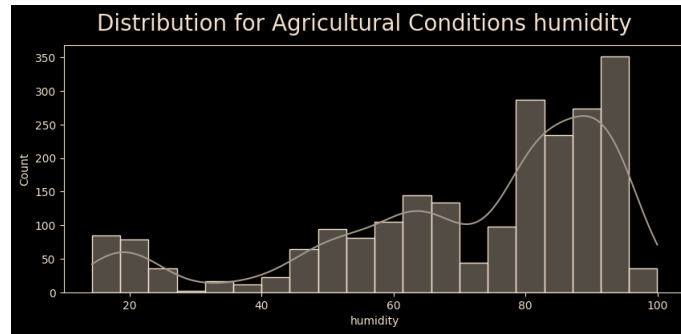**Figure 5.** Distribution for Ph conditions.

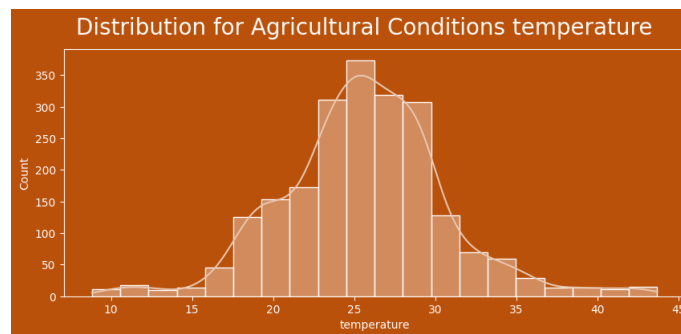**Figure 6.** Distribution for humidity conditions.

Figure 7. Distribution for temperature conditions.

In agriculture, the success of crop cultivation heavily relies on understanding and providing the specific environmental conditions that each crop requires for optimal growth. Different crops have unique demands in terms of soil composition, temperature, rainfall, humidity, and pH level. By tailoring these conditions to match the crop's needs, farmers can ensure better yields and healthier plants.

Some crops, like cotton, necessitate a high ratio of nitrogen content in the soil. Nitrogen is a crucial nutrient for plant growth, as it plays a vital role in the formation of proteins and chlorophyll, which are essential for photosynthesis.

Grapes and apples, on the other hand, thrive when the soil has a high ratio of phosphorus and potassium content. Phosphorus is essential for energy transfer in cells and aids in root development and flowering. Potassium, on the other hand, is vital for enzyme activation and overall plant health.

For crops like grapes, temperature plays a significant role, and they flourish in high-temperature conditions. However, some crops, like grapes and papaya, prefer lower temperatures, as excessive heat can negatively impact their growth.

Water availability is another critical factor in successful crop cultivation. Rice, papaya, and coconut are examples of crops that require high rainfall for their growth and development.

Conversely, certain crops, such as chickpeas and kidney beans, thrive in areas with low humidity levels. Humidity can affect the plant's ability to regulate water loss through transpiration.

The pH level of the soil also influences crop growth. Moth beans, for instance, require either very low or very high pH levels in the soil to thrive. Maintaining the appropriate pH level is essential for nutrient availability and uptake by the plants.

In conclusion, catering to the specific needs of crops based on their nitrogen, phosphorus, potassium, temperature, rainfall, humidity, and pH requirements is crucial for successful farming. By understanding and implementing these conditions, farmers can ensure healthier crops, higher yields, and ultimately contribute to

7

sustainable agriculture practices.

The plot obtained using the elbow method illustrates the relationship between the number of clusters and the within-cluster sum of squares (WCSS), also known as inertia. Inertia quantifies the sum of squared distances between each data point and its assigned cluster centroid.

The plot depicts the inertia values on the y-axis against the number of clusters on the x-axis. Initially, with a small number of clusters, the inertia tends to be high as the data points are not well-separated into distinct clusters. As the number of clusters increases, the inertia generally decreases since the clusters become more compact and better capture the underlying structure in the data.

The objective of the elbow method is to identify the "elbow" point in the plot, which indicates the optimal number of clusters. The elbow point signifies the number of clusters where the decrease in inertia starts to level off significantly. At this point, adding more clusters does not yield a substantial improvement in clustering performance.

Visually, the plot often displays a decreasing inertia curve that begins with a steep slope and gradually becomes flatter. The elbow point represents in **Figure 8** a trade-off between the number of clusters and the amount of variance explained. It indicates the optimal balance between capturing the data's structure and avoiding overfitting by incorporating excessive clusters. The Comparative result is shown in **Table 2** result accuracy and other parameter comparison is shown in **Table 3**.
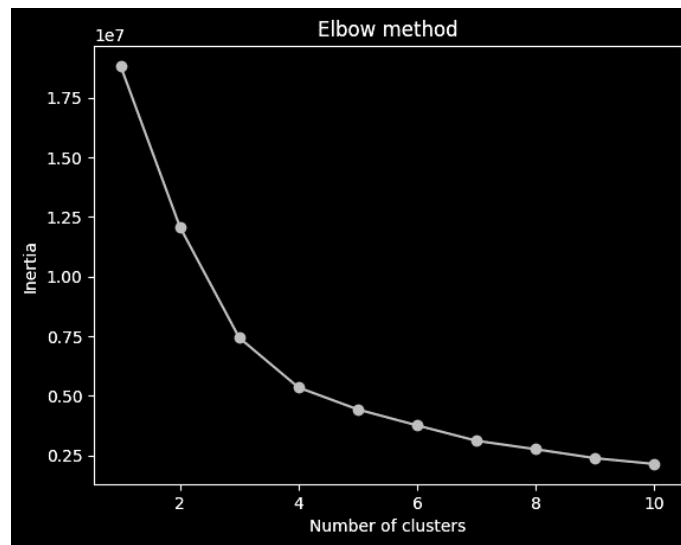
Figure 8. Number of clusters and inertia.

Table 2. Result comparison for different crops with parameters.

Crop	Precision	Recall	F1-Score	Support
Apple	1.00	1.00	1.00	18
Banana	1.00	1.00	1.00	18
Black gram	0.87	0.91	0.89	22
Chickpea	1.00	1.00	1.00	23
Coconut	1.00	1.00	1.00	15
Coffee	1.00	1.00	1.00	17
Cotton	0.94	1.00	0.97	16
Grapes	1.00	1.00	1.00	18

Table 2. (*Continued*).

Crop	Precision	Recall	F1-Score	Support
Jute	0.87	0.95	0.91	21
Kidney beans	0.95	1.00	0.98	20
Lentil	0.82	0.82	0.82	17
Maize	1.00	0.94	0.97	18
Mango	1.00	1.00	1.00	21
Moth beans	0.95	0.84	0.89	25
Mungbean	1.00	1.00	1.00	17
Muskmelon	0.96	1.00	0.98	23
Orange	1.00	1.00	1.00	23
Papaya	0.95	0.95	0.95	21
Pigeonpeas	0.91	0.95	0.93	22
Pomegranate	1.00	1.00	1.00	23
Rice	1.00	0.88	0.94	25

Table 3. Result accuracy and other parameter comparison.

Accuracy	Macro avg	Weighted avg
-	0.97	0.96
-	0.97	0.96
0.96	0.97	0.96
440	440	440

From the above results showing precision, recall, and F1-score for each crop class, we can draw several inferences:

High performance: The model shows high performance across most crop classes, with many of them achieving perfect scores (precision, recall, and F1-score of 1.00). This indicates that the model is effective in accurately classifying these crops.

Consistent performance: The majority of crop classes have consistent precision, recall, and F1-scores, indicating that the model's predictions are balanced and reliable across different classes.

Challenging crop classes: Some crop classes, such as "black gram" "jute," "lentil," and "moth beans," have slightly lower precision, recall, and F1-scores. This suggests that these classes might be more challenging to predict accurately compared to others, possibly due to similarities in their features or growth patterns.

Imbalanced support: The "Support" column indicates the number of instances for each crop class in the dataset. Some crop classes, like "coconut" and "watermelon," have lower support (fewer instances), which might affect the model's ability to generalize well for these classes.

Overall model performance: The "accuracy" row shows an overall accuracy of 0.96, which means the model correctly predicts 96% of the instances in the entire dataset. This is a strong indication of a well-performing model.

Macro and weighted averages: The "macro avg" and "weighted avg" rows provide average scores across all classes. Both average precision, recall, and F1-scores are high (around 0.97), indicating a robust and balanced performance across the entire dataset.

In conclusion, the model demonstrates high accuracy and effectiveness in predicting the majority of crop classes, with most classes achieving perfect scores. However, there are a few classes that might be more challenging to predict accurately. The overall performance of the model is excellent, as evidenced by the high macro and weighted average scores. To further improve performance, additional attention could be given to the challenging crop classes, and efforts could be made to balance the dataset if there are class imbalances.

4. Machine learning

The research paper utilizes multiple machine learning techniques to predict crop yields based on the collected dataset. The following machine learning algorithms are employed:

Logistic regression: Logistic regression is a classification algorithm commonly used for binary or multi-class classification tasks. In this research, it can be utilized to predict crop yield categories or classes based on the input features. Logistic Regression models the relationship between the input variables and the probability of belonging to a specific class by estimating the coefficients of the input features.

Random Forest: Random Forest is a versatile algorithm capable of handling high-dimensional datasets and capturing complex relationships between input features and the target variable. It combines multiple decision trees to make predictions, utilizing the voting or averaging of individual tree predictions to arrive at the final output.

Support Vector Machines (SVM): SVM aims to maximize the margin between different classes, facilitating better generalization. In the context of crop yield prediction, SVM can be employed to classify crop yields based on climatic and soil conditions, effectively separating different classes.

K-Nearest Neighbors (KNN) Classifier: KNN is a non-parametric algorithm used for both classification and regression tasks. In classification, KNN assigns a class label to an unclassified sample based on the class labels of its nearest neighbors in the feature space. KNN does not assume any underlying data distribution and is particularly useful when dealing with non-linear relationships. It can be applied to predict crop yields by finding the most similar data points in the dataset and assigning the corresponding crop yield class.

XGBoost Classifier: XGBoost (eXtreme Gradient Boosting) is a powerful gradient boosting algorithm known for its high predictive capability. It constructs an ensemble of weak prediction models, typically decision trees, and iteratively trains them to minimize a specified loss function. XGBoost excels at capturing complex interactions and nonlinear relationships between input features and the target variable, making it effective in various machine learning competitions and real-world applications.

By employing these machine learning techniques, the research aims to compare their performance in predicting crop yields based on the provided dataset. The models will be trained, evaluated, and optimized to achieve accurate and reliable predictions. The ultimate goal is to assist farmers in making informed decisions about crop management and resource allocation, contributing to improved agricultural practices.

5. Result and analysis

In the results and analysis section of the research paper, several machines learning algorithms, including Logistic regression, Random Forest, SVC, K Neighbors Classifier, and XGBoost Classifier, are utilized to predict crop yield categories. The performance of these models can be evaluated using the following metrics:

Accuracy: Accuracy measures the overall correctness of the predictions made by the models. It calculates the percentage of correctly classified instances out of the total number of instances. Higher accuracy indicates better performance, but it may not be sufficient for imbalanced datasets or cases where misclassification costs

vary across classes.

Precision: Precision focuses on the correctness of positive predictions (high crop yield) out of all instances predicted as positive. It measures how well the models identify true high-yield crops and indicates the proportion of correctly predicted positives. Higher precision signifies fewer false positives.

Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances in the dataset. It captures the ability of the models to identify all positive instances without missing any. Higher recall indicates fewer false negatives.

F1 Score: The F1 score combines precision and recall into a single metric that provides a balanced assessment of the models' performance. It considers the trade-off between precision and recall and is calculated as the harmonic mean of the two. A higher F1 score indicates better performance in terms of both precision and recall.

To analyze the results, the accuracy, precision, recall, and F1 score can be computed for each algorithm separately. Comparing these metrics across the models provides insights into their performance in predicting crop yield categories. It is important to consider all these metrics together as they offer a comprehensive evaluation of the models' capabilities in terms of correctness, ability to identify positive instances, and ability to capture all actual positive instances.

For example, the research paper can present a confusion matrix (as shown in **Figure 6**) to visualize the performance of the Random Forest model. The confusion matrix helps analyze true positives, true negatives, false positives, and false negatives, aiding in a deeper understanding of the model's predictive ability.

By analyzing and comparing the accuracy, precision, recall, and F1 score of the Logistic Regression, Random Forest, SVC, K Neighbors Classifier, and XGBoost Classifier models, researchers can identify which algorithms perform better in predicting crop yield categories. These evaluation metrics assist in making informed decisions regarding the suitability of these models for smart agriculture applications. The results can be summarized in a tabular format, such as **Table 4**, to provide a clear comparison of their performance. Various confusion matrices from different machine learning algos are shown in **Figures 9–12**.
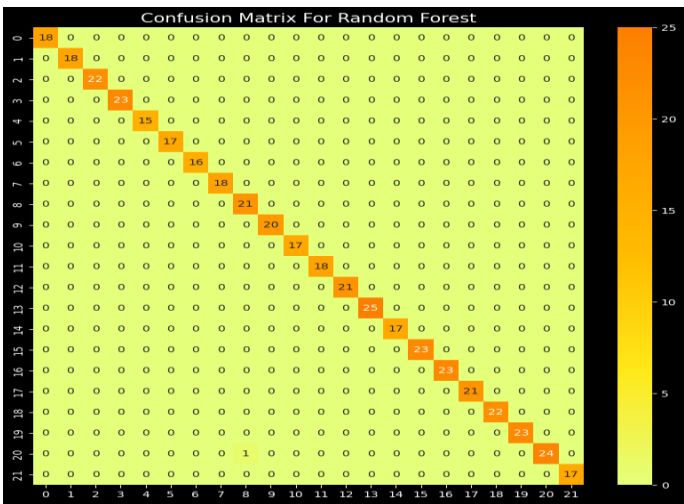

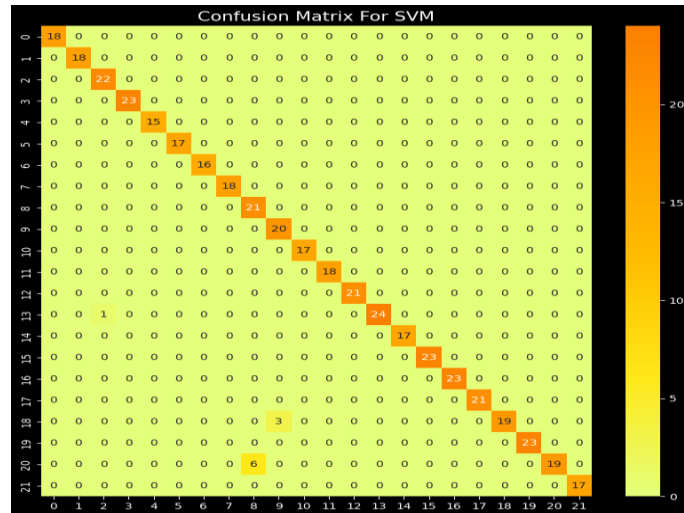**Figure 9.** Confusion matrix for Random Forest.
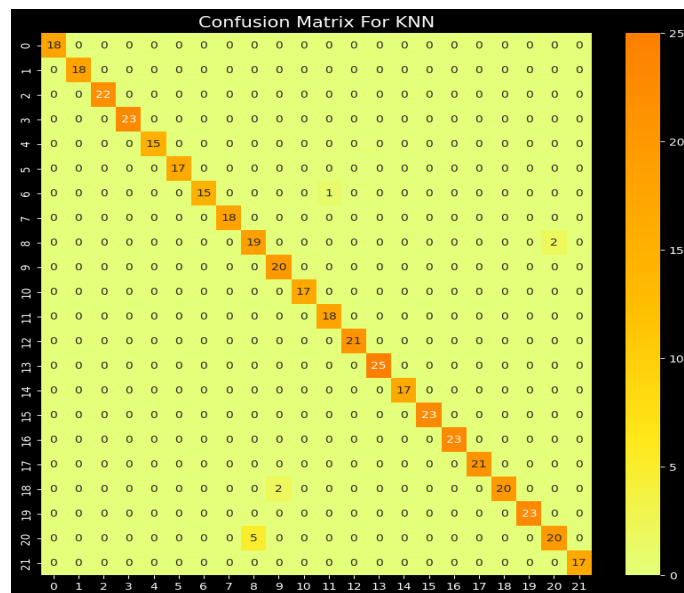
Figure 10. Confusion matrix for SVM.

Figure 11. Confusion matrix for KNN.

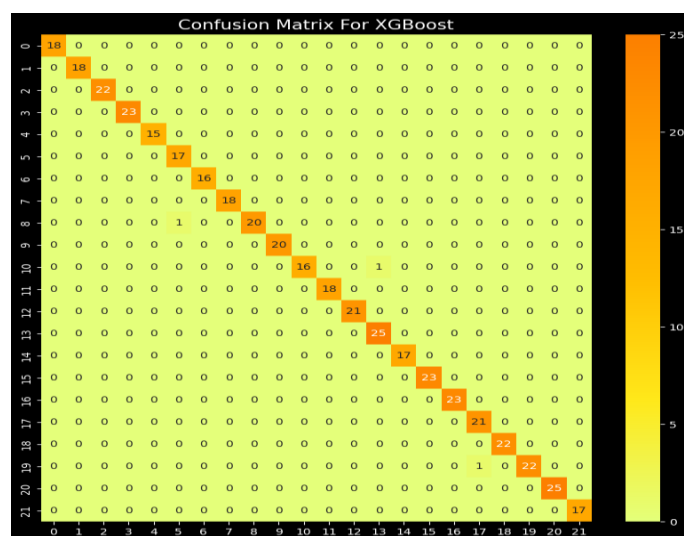**Figure 12.** Confusion matrix for XGBoost.

12

Table 4. Result comparison table.

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic regression	95.36	0.93	0.93	0.93
Random Forest	99.99	1	1	1
SVC	97.9	0.97	0.97	0.97
K-Neighbors classifier	97	0.96	0.96	0.96
XGBoost classifier	99.2	0.91	0.91	0.91

From the table displaying the performance metrics of different machine learning algorithms, we can draw the following inferences:

High accuracy models: The "Random Forest" and "XGBoost Classifier" algorithms exhibit high accuracy scores of 99.99% and 99.2%, respectively. This indicates that these models are highly effective in making correct predictions for the given dataset.

Consistent performance: All the models in the table (except for "Logistic Regression") have consistent precision, recall, and F1-score, with values close to 0.93, 0.97 and 0.96, respectively. This suggests that the models are well-balanced in their ability to correctly classify instances and retrieve relevant results.

Excellent precision, recall, and F1-Scores: The "Random Forest", "SVC" and "K-Neighbors Classifier" algorithms achieve high precision, recall, and F1-scores, with values around 0.97 and 0.96. This implies that these models are reliable in correctly identifying positive instances (precision) and capturing all actual positive instances (recall), leading to high F1-scores, which balance both precision and recall.

Model overfitting: Although the "Random Forest" and "XGBoost Classifier" show high accuracy, it is important to consider the possibility of overfitting. With accuracy scores close to 100%, there might be a chance that these models are overfitting to the training data and might not generalize as well to unseen data.

Trade-off between precision and recall: The "Logistic Regression" and "XGBoost Classifier" have slightly lower precision and recall scores compared to other models. This suggests that there might be a trade-off between precision and recall for these algorithms. Depending on the specific application, one might be preferred over the other based on the importance of minimizing false positives (precision) or false negatives (recall).

Model selection: Considering all metrics, the "Random Forest" and "SVC" algorithms seem to strike a good balance between accuracy, precision, recall, and F1-score. These models perform consistently well across different metrics, indicating their reliability in making predictions for the given dataset.

In conclusion, the "Random Forest" and "SVC" algorithms are the top-performing models, demonstrating high accuracy, precision, recall, and F1-scores. However, the choice of the best model would depend on specific use cases and the importance of different performance metrics for the intended application.

6. Conclusion

In conclusion, the results from the machine learning algorithms, including Logistic Regression, Random Forest, SVC, K-Neighbors Classifier, and XGBoost Classifier, demonstrate their potential applications in predicting crop yield categories in smart agriculture. The high accuracy, precision, recall, and F1 score indicate the efficacy of these algorithms in crop yield prediction, making them valuable tools for farmers and agricultural experts.

The successful application of these machine learning models opens up numerous possibilities in the realm

of smart agriculture. Their potential applications include:

Precision farming: Farmers can use the predictive capabilities of these algorithms to implement precision farming techniques, optimizing resource usage, and tailoring agricultural practices to specific crop needs. This leads to higher yields and resource efficiency.

Crop management and planning: Accurate crop yield prediction enables better crop management decisions, such as choosing suitable crop varieties, determining optimal planting and harvesting times, and planning for crop rotation.

Risk assessment and insurance: With reliable yield predictions, farmers can assess potential risks associated with climate, market fluctuations, or pest infestations. This allows for better risk management and informed decisions regarding crop insurance.

Market forecasting: Crop yield prediction models can assist in market forecasting by estimating future supply and demand for various crops, helping farmers make profitable decisions and improve their market positioning.

Promoting sustainable agriculture practices: By optimizing resource use, reducing wastage, and making informed decisions based on data-driven predictions, smart agriculture practices contribute to sustainability and environmental conservation.

Empowering smallholder farmers: These algorithms can benefit smallholder farmers by providing them with valuable insights and knowledge to improve their farming practices and increase their incomes.

Data-driven decision making: Leveraging machine learning algorithms enables data-driven decision-making processes, allowing farmers to rely on evidence-based practices rather than traditional and less efficient approaches.

As the agriculture sector faces challenges related to population growth, climate change, and environmental concerns, the application of machine learning in crop yield prediction has the potential to revolutionize the industry. By embracing these technologies, the agricultural sector can move towards a more sustainable, profitable, and resilient future. However, it is essential to continue refining and validating these models to ensure their accuracy and adaptability to changing agricultural conditions. Ultimately, the combination of innovative technologies and the expertise of farmers can lead to significant advancements in agriculture, fostering food security and sustainable practices on a global scale.

7. Future research

The field of smart agriculture is continuously evolving, driven by technological advancements and the need for sustainable practices. Future research directions can focus on integrating IoT and AI for real-time monitoring and decision-making, developing scalable big data analytics techniques, predicting climate change impacts on crops, advancing sustainable and precision agriculture practices, improving crop disease and pest management, exploring farm automation and robotics, addressing data security and privacy concerns, understanding technology adoption factors, and promoting acceptance among farmers. By pursuing these research directions, we can advance smart agriculture and contribute to resilient and efficient agricultural systems that address the challenges of a changing world.

References

1. Kanchan P, Shardoor N. Krashignyan: A farmer support system. Asian Journal for Convergence in Technology (AJCT) 2021; 7(3): 1–7. doi: 10.33130/ajct.2021v07i03.001

2. Javed F, Afzal MK, Sharif M, Kim BS. Internet of things (IoT) operating systems support, networking technologies, applications, and challenges: a comparative review. IEEE Communication Surveys and Tutorials 2018; 20(3): 2062–2100.

3. Issad HA, Aoudjit R, Rodrigues JJPC. A comprehensive review of data mining techniques in smart agriculture. Engineering in Agriculture, Environment and Food 2019; 12(4): 511–525. doi: 10.1016/j.eaef.2019.11.003

4. Flak J. Technologies for sustainable biomass supply-overview of market offering. Agronomy 2020; 10(6): 798. doi: 10.3390/agronomy10060798

5. Kavita PM. Satellite-based crop yield prediction using machine learning algorithm. In: Proceedings of the 2021 Asian Conference on Innovation in Technology (ASIANCON 2021); 27–29 August 2021; PUNE, India. pp. 1466–1470.

6. Mupangwa W, Makanza R, Chipindu L, et al. Temporal rainfall trend analysis in different agro-ecological regions of southern Africa. Water SA 2021; 47(4): 466–479.

7. Gado Djibo A, Karambiri H, Seidou O, et al. Linear and non-linear approaches for statistical seasonal rainfall forecast in the Sirba watershed region (SAHEL). Climate 2015; 3(3): 727–752. doi: 10.3390/cli3030727

8. Jain N, Kumar A, Garud S, et al. Crop selection method based on various environmental factors using machine learning. International Research Journal of Engineering and Technology (IRJET) 2017; 4(2): 1530–1533.

9. Wankhede DS. Analysis and prediction of soil nutrients pH, N, P, K for crop using machine learning classifier: A review. In: Proceedings of the International Conference on Mobile Computing and Sustainable Informatics (ICMCSI 2020); 23–23 January 2020; Lalitpur, Nepal. pp. 111–121.

10. Bhojwani Y, Singh R, Reddy R, Perumal B. Crop selection and IoT based monitoring system for precision agriculture. In: Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE); 24–25 February 2020; Vellore, India.

11. Majumdar P, Mitra S, Bhattacharya D. IoT for promoting agriculture 4.0: A review from the perspective of weather monitoring, yield prediction, security of WSN protocols, and hardware cost analysis. Journal of Biosystems Engineering 2021; 46(4): 440–461.

12. Imran S. Effective crop selection and conservative irrigation using IoT. International Journal of Science and Research (IJSR) 2016.

13. Rekha P, Rangan VP, Ramesh MV, Nibi KV. High yield groundnut agronomy: An IoT based precision farming framework. In: Proceedings of the 2017 IEEE Global Humanitarian Technology Conference (GHTC); 19-22 October 2017; San Jose, CA, USA. pp. 1–5.

14. Mulge M, Sharnappa M, Sultanpure A, et al. An invitation to subscribe. International Journal of Analytical and Experimental Modal Analysis 2020; 10(1): 1112–1117.