IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13801

# Performance Evaluation of Machine Learning Algorithms in Smart Agriculture

**Dennis Karugu Gichuki[1], Patrick Oduor Owoche[2], Samuel Mungai Mbuguah[3]**

PhD Candidate, Department of Information Technology, Kibabii University, Bungoma, Kenya[1]

Lecturer, Department of Information Technology, Kibabii University, Bungoma, Kenya[2]

Associate Professor, Department of Information Technology, Kibabii University, Bungoma, Kenya [3]

**Abstract**: This study explores the integration of Wireless Sensor Networks (WSN) and Machine Learning (ML) in smart farming to address critical agricultural challenges. By leveraging real-time data collection and advanced analytical tools, the research demonstrates the potential of ML algorithms—Decision Trees, Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forests—in enhancing crop management, including yield prediction, soil quality assessment, and pest and disease detection. The study finds that Naive Bayes achieves the highest accuracy and balanced precision-recall metrics, while ensemble methods like Random Forests effectively reduce overfitting and improve prediction accuracy. Despite the promising results, the research identifies challenges such as data accessibility, model integration, and user interface design that must be addressed to fully realize the potential of smart farming technologies. Overall, the findings provide valuable insights into optimizing resource utilization, reducing crop losses, and promoting sustainable farming practices, thereby supporting global food security and economic stability.

**Keywords:** Smart Farming, Machine Learning, Supervised Learning, Data Drive Decision

## I. INTRODUCTION

Crops are faced with several factors throughout their lifecycle such as environmental factors which include temperature, soil, moisture, physical, pests and many more [1]. These tend to affect the crop's yield as well as crop's quality which continue to threaten food security as well as contribute huge losses to the global economy. These challenges have continued to contribute to a significant decline in agricultural produce which consequently threatens food security in the face of the ever-increasing world population. The situation is however not helpless as smart farming technologies promise efficiency as well as enhanced productivity as compared to conventional agricultural approaches [2].

Smart farming, a concept used to refer to the application of digital technologies that interact through networks in agriculture has the potential to make agriculture more profitable as it reduces the cost of production [3]. The main aim of smart farming is to replace traditional agriculture with advanced technologies to enhance sustainability and productivity by employing efficient tools [4]. Unlike precision agriculture which only takes into account in-field variability, smart farming goes beyond basing management tasks not just on location but also on context enhanced data and situation awareness that is triggered by real-time events. Unlike conventional approaches, smart farming technologies look into the problem of what, when and where resources should be applied for maximum yield [5].

Further, prediction is possible by gathering and analyzing farming factors including soil parameters, fertilizer requirements as well as weather which are essential for optimum crops performance [6]. The accuracy of predictions largely depends on the quantity and quality of available data as well as the proper application of machine learning algorithms (Haque et al. 2021). Presently, diverse machine learning approaches ranging from supervised learning, unsupervised learning and reinforcement learning have been proposed to tackle the variability in crops, soils, fertilizer as well as climate suitability [7].

## II. STATEMENT OF THE PROBLEM

The agricultural sector is pivotal to the sustenance of human life and the global economy, yet it faces numerous challenges that threaten productivity and sustainability. Traditional farming practices are often inefficient and labour-intensive, leading to suboptimal crop yields and resource utilization [8]. Furthermore, the increasing frequency of extreme weather events, pests, and diseases exacerbates the unpredictability of agricultural outputs.

DOI: 10.17148/IJARCCE.2024.13801

One of the critical issues in modern agriculture is the timely and accurate detection of plant health problems, such as pest infestations and diseases. Conventional methods of monitoring and diagnosing these issues are predominantly manual, time-consuming, and prone to human error, resulting in delayed interventions and significant crop losses. Additionally, the lack of precise data-driven insights hinders farmers from making informed decisions that optimize resource use and improve crop management.

In recent years, advancements in technology have opened new avenues for addressing these challenges. Wireless Sensor Networks (WSN) provide real-time data on various environmental parameters, while Machine Learning (ML) algorithms can analyze this data to detect patterns and make predictions. However, despite the potential of ML in transforming agricultural practices, its application remains limited due to several factors, including the lack of accessible and accurate datasets, the complexity of integrating WSN data with ML models, and the need for user-friendly interfaces that can be readily adopted by farmers.

Therefore, this study aims to explore the integration of WSN and ML in agriculture, focusing on developing a high-performance model for the detection of plant pests and diseases. By leveraging model ensemble techniques, including the Inception module and cluster algorithms, the research seeks to enhance the accuracy and reliability of plant health monitoring systems. This approach promises to provide farmers with timely, actionable insights, thereby improving crop yields, reducing resource wastage, and promoting sustainable farming practices.

## III. LITERATURE REVIEW

In the context of smart farming, various machine learning algorithms such as Decision Trees, Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forests have been extensively studied and applied. These algorithms play a crucial role in addressing diverse agricultural challenges, ranging from crop yield prediction to pest and disease detection [8].

Decision Trees and Random Forests are particularly effective for modelling complex relationships in agricultural data. Their application in tasks such as crop yield prediction and soil quality assessment has demonstrated significant potential. Random Forests, with their ensemble learning approach, provide robust performance by reducing overfitting and handling large datasets efficiently. This makes them a valuable tool for farmers aiming to optimize crop management and improve yield predictions [9].

Naive Bayes classifiers, known for their simplicity and efficiency, are suitable for real-time applications in smart farming. They have been successfully used for classifying crop diseases based on image data, enabling timely alerts for farmers to take preventive measures. The algorithm's ability to handle noisy data and provide probabilistic outputs makes it a practical choice for disease prediction and pest detection in crops [8].

Support Vector Machines (SVMs) are favoured in precision agriculture for tasks such as crop type classification and weed detection. Their capacity to handle high-dimensional data and perform well with limited samples is particularly beneficial in agricultural settings where data collection can be expensive or challenging. Comparative studies have shown that SVMs often outperform other methods in terms of classification accuracy, making them a reliable choice for precision farming applications [11].

Logistic Regression is commonly applied to binary classification problems in agriculture, such as predicting the presence or absence of pests or diseases [12]. Its interpretability and ease of implementation are significant advantages, making it a popular choice for developing early warning systems. When combined with spatial data, Logistic Regression has proven effective in predicting areas at risk of pest infestations, thus aiding in targeted pest management strategies [11].

Comparative studies evaluating these algorithms in various agricultural applications highlight their relative strengths. For instance, one study assessed the performance of Decision Trees, Random Forests, SVM, Naive Bayes, and Logistic Regression in predicting landslide susceptibility, a problem analogous to certain agricultural risk assessments. The findings indicated that ensemble methods like Random Forests often provided superior performance due to their robustness and accuracy in prediction tasks [12] [13].

These empirical studies underscore the versatility and effectiveness of machine learning algorithms in enhancing smart farming practices. By leveraging these advanced analytical tools, farmers can make more informed decisions, optimize resource use, and ultimately improve agricultural productivity and sustainability.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13801

## IV. RESULTS AND DISCUSSION

To evaluate machine learning algorithms' performance in the smart farming domain, this section presents data on various performance metrics tailored to the specific nature of the recommendation algorithm and the type of problem it addresses, such as classification, regression, or ranking. For classification tasks, where the recommender system predicts the optimal crop to grow based on given input features, several key metrics are employed. These include Accuracy, Precision, Recall, and the F1 Score.

TABLE 1: ML ALGORITHMS PERFORMANCE METRICS

| Metric | Decision Trees | Naive Bayes | SVM | Logistic Regression | Random Forest |
|---|---|---|---|---|---|
| Accuracy | 0.868 | 0.937 | 0.930 | 0.901 | 0.914 |
| Macro Average Precision | 0.850 | 0.941 | 0.926 | 0.903 | 0.917 |
| Macro Average Recall | 0.865 | 0.941 | 0.939 | 0.905 | 0.922 |
| Macro Average F1-Score | 0.854 | 0.941 | 0.930 | 0.901 | 0.917 |
| Weighted Average Precision | 0.841 | 0.940 | 0.931 | 0.897 | 0.912 |
| Weighted Average Recall | 0.871 | 0.941 | 0.933 | 0.901 | 0.916 |
| Weighted Average F1-Score | 0.849 | 0.940 | 0.930 | 0.900 | 0.912 |

The findings presented in Table 1 indicate that the accuracy of a model reflects its overall performance in correctly classifying instances across all classes. Among the models evaluated, Naive Bayes achieved the highest accuracy at 93.7%. This indicates that Naive Bayes consistently performs well in identifying the correct class for a given instance. The Support Vector Machine (SVM) followed closely with an accuracy of 93.0%, demonstrating its robust classification capability. Random Forest (RF) came next with an accuracy of 91.4%, showcasing its high performance. Logistic Regression recorded an accuracy of 90.1%, which is strong but slightly lower compared to Naive Bayes and SVM. Decision Trees had the lowest accuracy at 86.8%, suggesting comparatively weaker performance in correctly classifying instances.

Macro average precision evaluates the precision of the model by averaging the precision scores for each class without considering class imbalance. Naive Bayes achieved the highest macro average precision at 94.1%, indicating it provides the most accurate classifications across different classes. Random Forest followed with a macro average precision of 91.7%, reflecting its strong performance in precision. The Support Vector Machine (SVM) had a macro average precision of 92.6%, indicating good precision. Logistic Regression had a macro average precision of 90.3%, demonstrating solid performance but lower than Naive Bayes and SVM. Decision Trees had the lowest macro average precision at 85.0%, highlighting its relatively weaker performance in this metric.

Macro average recall measures the model's ability to identify all relevant instances for each class and then averages the recall scores. Naive Bayes achieved the highest macro average recall at 94.1%, reflecting its excellent ability to detect relevant instances across classes. Random Forest had a macro average recall of 92.2%, showing strong performance in the recall. The Support Vector Machine (SVM) followed with a macro average recall of 93.9%, demonstrating robust recall capability. Logistic Regression recorded a macro average recall of 90.5%, indicating effective performance but slightly lower than Naive Bayes and SVM. Decision Trees had the lowest macro average recall at 86.5%, indicating comparatively weaker performance in this aspect.

The macro average F1-score combines precision and recall into a single metric by averaging the F1-scores of all classes. Naive Bayes achieved the highest macro average F1-score at 94.1%, representing the best balance between precision and recall. Random Forest followed with a macro average F1-score of 91.7%, indicating strong overall performance. The Support Vector Machine (SVM) had a macro average F1-score of 93.0%, reflecting a good balance between precision and recall. Logistic Regression recorded a macro average F1-score of 90.1%, showing solid performance. Decision Trees had the lowest macro average F1-score at 85.4%, indicating a weaker balance compared to other models.

Weighted average precision accounts for class imbalance by averaging the precision scores of each class, weighted by the number of instances in each class. Naive Bayes achieved the highest weighted average precision at 94.0%, demonstrating the best overall precision across classes. Random Forest followed with a weighted average precision of 91.2%, showing high precision. The Support Vector Machine (SVM) had a weighted average precision of 93.1%, indicating strong performance.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13801

Logistic Regression recorded a weighted average precision of 89.7%, reflecting effective performance. Decision Trees had the lowest weighted average precision at 84.1%, suggesting relatively weaker performance. Weighted average recall measures the model's ability to correctly identify relevant instances, considering class imbalance. Naive Bayes achieved the highest weighted average recall at 94.1%, reflecting its superior ability to detect relevant instances across all classes. Random Forest had a weighted average recall of 91.6%, showing strong performance. The Support Vector Machine (SVM) followed with a weighted average recall of 93.3%, demonstrating robust recall. Logistic Regression recorded a weighted average recall of 90.1%, indicating solid performance. Decision Trees had the lowest weighted average recall at 87.1%, showing comparatively weaker performance in identifying relevant instances.

The weighted average F1-score combines precision and recall, accounting for class imbalance, and averages these scores based on class distribution. Naive Bayes achieved the highest weighted average F1 score at 94.0%, reflecting the best overall balance between precision and recall. Random Forest followed with a weighted average F1-score of 91.2%, indicating strong performance. The Support Vector Machine (SVM) had a weighted average F1-score of 93.0%, demonstrating good balance. Logistic Regression recorded a weighted average F1-score of 90.0%, showing effective performance. Decision Trees had the lowest weighted average F1-score at 84.9%, indicating a relatively weaker balance. The performance evaluation of various machine learning models on the dataset is summarized in Table 1 while accuracy comparison is shown in Figure 1. The accuracy of each model is reported to four decimal places for precision.

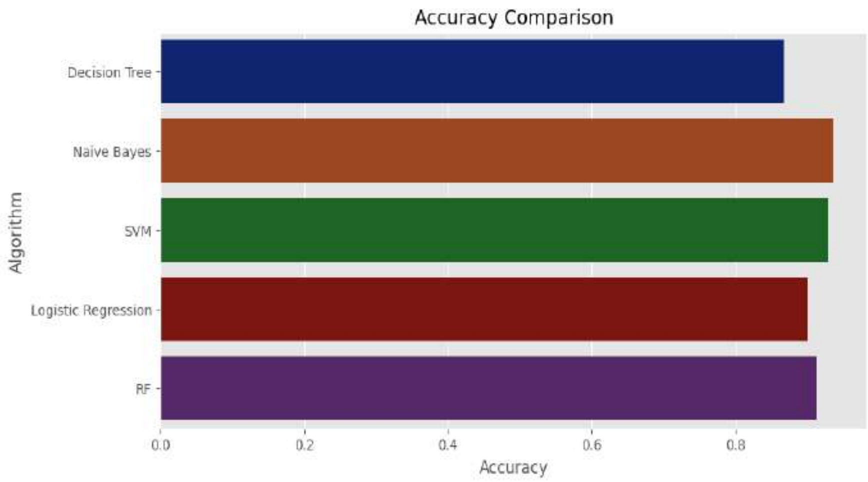


Figure. 1 Accuracy Comparison

## V. CONCLUSION

In conclusion, integrating Wireless Sensor Networks (WSN) and Machine Learning (ML) in smart farming presents a transformative approach to addressing significant agricultural challenges. This study demonstrates the effectiveness of various ML algorithms, including Decision Trees, Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forests, in enhancing crop management through improved yield prediction, soil quality assessment, and pest and disease detection. Naive Bayes, in particular, showed the highest accuracy and balanced precision-recall metrics, while ensemble methods like Random Forests excelled in reducing overfitting and improving prediction accuracy. Despite these advancements, challenges such as data accessibility, model integration, and user interface design must be addressed to fully realize the potential of smart farming technologies. Overall, the findings contribute valuable insights into how advanced technological solutions can optimize resource utilization, reduce crop losses, and promote sustainable farming practices, thereby supporting global food security and economic stability.

## REFERENCES

[1] M. Li, S. Cheng, J. Cui, C. Li, Z. Li, C. Zhou and C. Lv, "High-Performance Plant Pest and Disease Detection Based on Model Ensemble with Inception Module and Cluster Algorithm," *Plants*, vol. 12, no. 200, 2023.

[2] E. S. Kwaghtyo and J. C. Eke, "Advances in smart farming: Potential and benefits," *Agricultural Systems,* vol. 3, no. 2, pp. 45-56, 2022.

[3] E. Collado, A. Fossatti and Y. Saez, "Smart farming: A potential solution towards a modern and sustainable agriculture in Panama," *AIMS Agriculture and Food*, vol. 4, no. 2, p. 266–284, 2018.

[4] Aqeel-ur-Rehman, Towards smart agriculture: An introduction. Smart agriculture: An approach towards better agriculture management, OMICS Group eBooks, 2015.

[5] G. Adamides, "The future of smart farming: Towards sustainability and productivity," *Journal of Agricultural Informatics*, vol. 11, no. 3, pp. 55-67, 2020.

[6] S. Fahad, A. A. Bajwa, U. Nazir, S. A. Anjum, A. Farooq and A. Zohaib, "Crop production under drought and heat stress: Plant responses and management options," *Frontiers in Plant Science*, vol. 8, p. 1147, 2017.

[7] M. Parikh, R. Yadav and K. Singh, "Machine learning applications in agriculture: A review," *Artificial Intelligence in Agriculture*, vol. 5, pp. 1-10, 2021.

[8] A. Abbas, H. Zhang and W. Yang, "Proximal sensing and machine learning for crop yield prediction," *Journal of Agricultural Engineering Research*, vol. 11, no. 3, pp. 87-96, 2020.

[9] L. Chen, Z. Wang and J. Chen, "Image-based crop disease classification using Naive Bayes," *International Journal of Agricultural Informatics and Technology*, vol. 8, no. 4, pp. 256-267, 2019.

[10] X. Chen, B. Wu, Y. Xu, Y. Zhang and J. Li, "Application of machine learning in crop yield prediction," *Journal of Agricultural Science and Technology*, vol. 12, no. 3, pp. 345-359, 2020.

[11] I. Colkesen, T. Kavzoglu and M. Alkan, "Support vector machines in precision agriculture," *Journal of Precision Agriculture*, vol. 7, no. 2, pp. 199-214, 2016.

[12] Sangeeta, "Environmental factors and crop yield estimation using machine learning," *Journal of Agricultural Science and Technology*, vol. 15, no. 2, pp. 234-245, 2020.

[13] X. Chen and Y. Zhou, "Logistic regression in predicting pest infestations in agriculture," *Agricultural Data Science Journal*, vol. 15, no. 1, pp. 102-118, 2023.

[14] D. T. Bui, B. Pradhan, O. Lofman, I. Revhaug and O. B. Dick, "Landslide susceptibility assessment using machine learning models," *Journal of Environmental Management*, vol. 164, pp. 123-135, 2016.

[15] Y. Chen, X. Zhang and L. Feng, "Comparative study of machine learning models for landslide susceptibility prediction," *Computers and Geosciences*, vol. 98, pp. 1-15, 2017.

## BIOGRAPHY

**Dennis Gichuki** is a PhD candidate in Information Technology at Kibabii University with a focus on the integration of Wireless Sensor Networks (WSN) and Machine Learning (ML) for smart farming technologies in Kenya. His research aims to enhance sustainable farming practices through the development of advanced machine-learning algorithms that utilize real-time data from WSNs. He is a member of the Association of Computing Practitioners-Kenya and the Internet Society. He has a keen interest in systems virtualization, ML, and Information and Communication Technology for Development (ICT4D). Committed to leveraging technology to address agricultural challenges, improve crop yields, and promote sustainable development.

**Dr. Patrick Oduor Owoche,** with a PhD in Information Technology from Kibabii University, is a distinguished academic and IT professional known for his contributions to educational leadership and technological innovation. He has played a key role in promoting international collaborations and enhancing Kibabii University's global footprint through strategic partnerships. His technical expertise includes web and multimedia technologies, instructional design, and cloud application development. Dr. Owoche's active participation in seminars and workshops, including those at Leibniz Universität Hannover and Stellenbosch University, along with his role as a Visiting Scholar at Chandigarh University, highlights his dedication to academic excellence. His research focuses on applying information technology in educational settings to improve teaching methodologies and infrastructure. Beyond academia, he contributes to professional bodies and policy development in technology-enabled learning and international education strategies, impacting global educational practices and innovations.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⵣ Peer-reviewed & Refereed journal ⵣ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13801

**Prof. Samwel Mungai Mbuguah** is an Associate Professor of Information Technology and the Director of Planning and Organization Performance at Kibabii University. He has previously held roles such as Director of Information and Communication Technology and Acting Dean of the School of Computing and Informatics at Kibabii University. With a PhD in Information Technology, Prof. Mbuguah has supervised seven PhD students and seventeen Master's students to graduation, assessed 54 postgraduate theses from various universities, and published 70 papers in refereed journals, books, and book chapters. He is a Chartered Engineer and an assessor for Chartered Engineer applicants for the Engineering Council UK. Additionally, he is a professional member of the Association of Computing Machinery, the British Computer Society, and the Association of Computing Professionals in Kenya.