iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.64928

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Optimizing Rice Yield Quality Through Advanced Machine Learning Techniques in Precision Farming

Gbah Gonto Jean Claude[1], Togola Molobaly Di Bebe[2], Yunusa Ishaq[3]
Department Computer Science, Nanjing University of Information Science and Technology

Abstract: ***The continuously rising global trend in demographics is a challenge that modern agriculture has to contend with to feed the world by producing food produced with scarce inputs. In response, Advanced emerging technologies like machine learning (ML) and the IoT have birthed intelligent farming. This paper gives a decision-making framework for crop management, particularly using machine learning models to identify rice quality. Apache Kafka and Spark are used with the system and are coupled with big data techniques for enhanced data streaming and processing. Real-time classification of rice quality using various ML algorithms, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Feedforward Neural Networks (FFNN) are used. This research shows that this methodology significantly improves the decision-making approach in precision agriculture as a viable framework for future agricultural systems. Lastly, suggestions for future development aret presented regarding incorporating environmental information and utilizing the models in wider scouting practices.***

Keywords: Smart farming, Decision Support System (DSS), Machine Learning (ML), Rice quality prediction, Real-Time data streaming

I. INTRODUCTION

It is estimated that the total human populace is expected to be 9.7 billion by 2050, and there is a need to double food production rates globally simultaneously; water and arable land are scarce [1]. Specifically, through the use of the Internet of Things (IoT), artificial intelligence (AI), and machine learning (ML), smart farming offers valid opportunities for the enhancement of agriculture through a methodology that involves the employment of real-time data gathering, analysis, and planning. Expert systems, such as the Decision Support Systems (DSS), are used to understand various inputs, including soil and weather, and significantly recommend the most appropriate planting, watering, and harvesting regimes[2], [3], [4]. New developments in ML have improved the DSS's effectiveness in the forecasting of yield/ quality of crops[5], [6]. In this respect, field crops like rice that are vulnerable to factors such as temperature and water supply a model that uses ML for predicting quality, and the predictive parameters include the chemical composition of the food item and the phases of the crop's growth[7]. This paper evaluates a real-time ML-based DSS for rice quality prediction using Apache Kafka and Spark to facilitate real-time decision-making [8].

II. RELATED WORK

In recent years, machine learning (ML) applications have transformed decision support systems (DSS) in agriculture, particularly through predictive models that support real-time decision-making for crop management [9]. ML algorithms are applied across tasks such as yield estimation, disease detection, and quality classification [10], with studies showing high accuracy for models like support vector machines (SVM), neural networks, and decision trees [11]. For instance, Liakos et al. and Zhi Hong Kok et al. provide comprehensive reviews emphasizing the impact of these models in precision farming [12], [13]. However, they lack real-time data integration for operational use Addressing this gap, our research integrates Apache Kafka and Spark for real-time processing, enabling immediate, large-scale data handling crucial for modern agriculture [14]. Scalability and robustness are vital for DSS, particularly crop quality predictions reliant on high-dimensional IoT data. Existing studies, such as Giusti's fuzzy DSS for irrigation, demonstrate effective agricultural data use but are limited to real-time, high-scale analysis [15]. By leveraging a cloud-based architecture for data storage and visualization, our approach scales with large datasets and offers a dynamic framework for informed, real-time decisions in precision farming [16].

©IJRASET: All Rights are Reserved

III. METHODOLOGY

Based on our research, the decision-making principles for intelligent farming were created, emphasizing rice quality on-the-fly prediction. The system combines and incorporates many technologies, such as machine learning (ML) algorithms, big data, and the cloud. The architecture framework involved is arranged so that big data retrieval and forecast are feasible and real-time and assist in timely strategic planning of crop management.

A. Data Collection and Preprocessing

The data used in the study were obtained from the Chinese National Germplasm Repository, which has 50,000 samples with 80 original attributes. In this study, some output measures, for example, plant height, panicle number, and starch, were chosen, and therefore, the dataset was shrunk to 36 attributes after data preprocessing. In case of missing values, imputation followed by ASCII conversion for compatibility with the rest of ML models was applied. Normalization was used for feature scaling, ensuring that every input variable was normalized for training.

B. Machine Learning Algorithms

Three ML algorithms were chosen for their suitability in classification tasks and established efficacy in agricultural studies: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Feed Forward Neural Network (FFNN). The first algorithm, KNN with the Euclidean distance, is relatively simple and more suitable for small datasets; the second algorithm, SVM with an RBF kernel, is considered to be highly effective for work in high-dimensional space; the third type of FFNN with three hidden layers was used to identify non-linear relationships at the cost of high computational parameters.

C. System Architecture

The IoT sensors are included in the DSS architecture, with Apache Kafka as a data streaming service, Apache Spark for real-time data processing, and Onesait as a cloud platform for data storage and visualization. For synchronous data migration, Kafka produced a highly reliable and effort-free solution, while for processing, Spark brought in mini-batch data processing for parallel, efficient operation. Onesait provided support for data storage and gave an announcement interface for live crop surveillance and quality checking.



Fig1. Spark streaming data flow diagram

D. Experimental Setup

The system was tested using historical rice data for model training, with live sensor data streamed through the Kafka-Spark pipeline. Model performance was assessed through accuracy, precision, and F1-score metrics, while latency and scalability tests confirmed system robustness under varying data loads.

IV. RESULTS

The KNN, SVM, and FFNN ML models were evaluated based on accuracy, precision, recall, F1 score, and computational cost. The key performance indicators and their definitions are outlined in Table 1.

TABLE I. PERFORMANCE METRICS FOR KNN, SVM, AND FFN MODELS

Model	Accuracy (%)	Precision	Recall	F1-Score	Computational efficiency (time)
KNN	68.5	0.67	0.68	0.67	medium
SVM	70.3	0.71	0.70	0.70	high
FFNN	72.8	0.73	0.73	0.73	low

A. Machine Learning Model Performance

The KNN model achieved 68.5% accuracy for rice quality classification, with precision and an F1-score of 0.67. They also found that increasing the degree of mock neighbors (k) influenced the accuracy of the program in finding the real neighbors, and k=5 was the optimum. However, misclassification was observed in the "low" and "high" quality classes. Accuracy is depicted in Figure 2, and computation times remain stable at various k values regardless of the configurations. This shows that KNN is best used in cases where a prioritization on computational speed outweighs a fractionally better accuracy result.

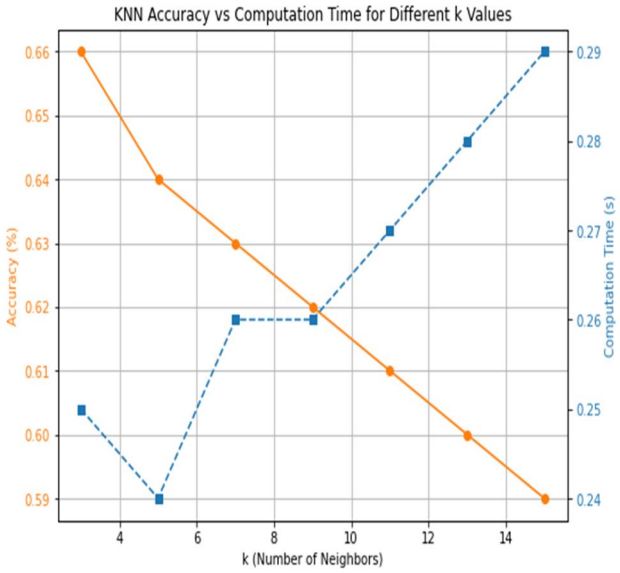


Fig2. KKK Accuracy vs Computation time for different k values

Unlike KNN, which scored only 59 %, the SVM model with a radial basis function using the RBF kernel attained a higher accuracy of 71%. The optimization of the penalty parameter (C) and kernel coefficient (gamma) significantly improved classification performance, particularly in separating the "high" and "low" quality classes. As illustrated in Figure 3, SVM's accuracy increased with larger training datasets, while computation time remained efficient due to the model's scalability.

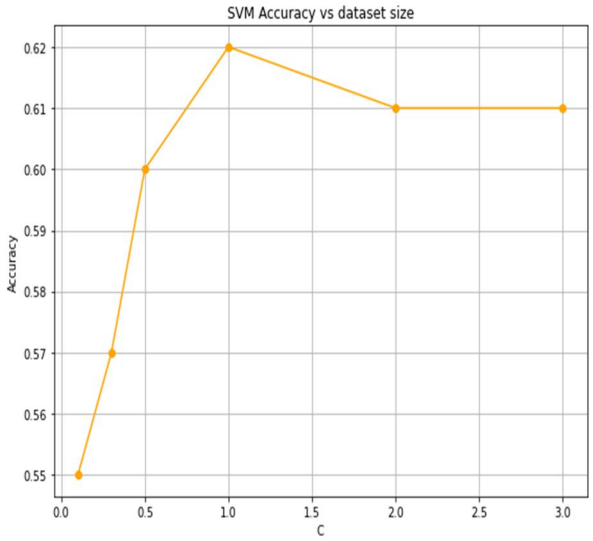


Fig3. SVM Accuracy vs Dataset

SVM outperformed KNN in accuracy and generalization, making it a preferred model for predicting rice quality in more extensive and complex datasets. The confusion matrix, shown in Figure 4, demonstrates that SVM performed well in predicting the "medium" rice quality class, but there was still some overlap between the "low" and "medium" classes.

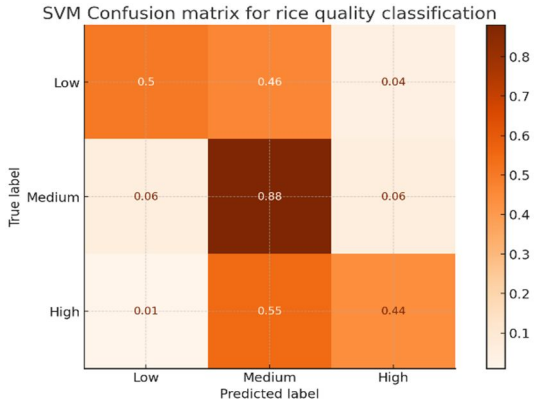


Fig4. Confusion matrix for rice quality classification

The FFN model yields the highest accuracy at 77%, with an F1-score of 0.73. three hidden-layer architectures facilitated complex relationship mapping, but the model's computational demands were substantial. Figure 5 highlights the accuracy versus computation time for FFNN, while the loss curve in Figure 5 shows the model's convergence over 1,000 iterations, affirming training effectiveness. Although FFNN achieved superior accuracy, its computational cost may limit real-time application.

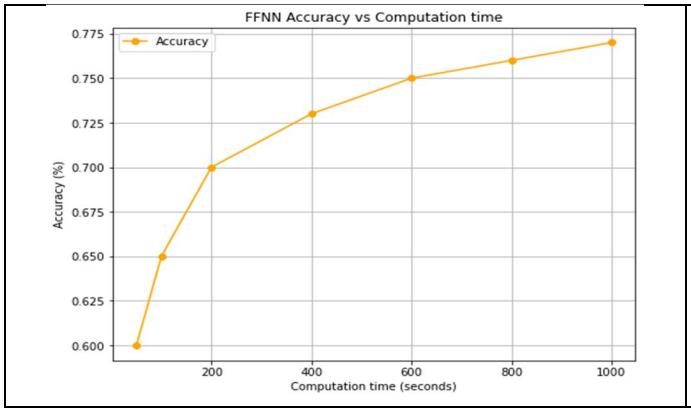


Fig5 FFNN Accuracy vs. Computation time

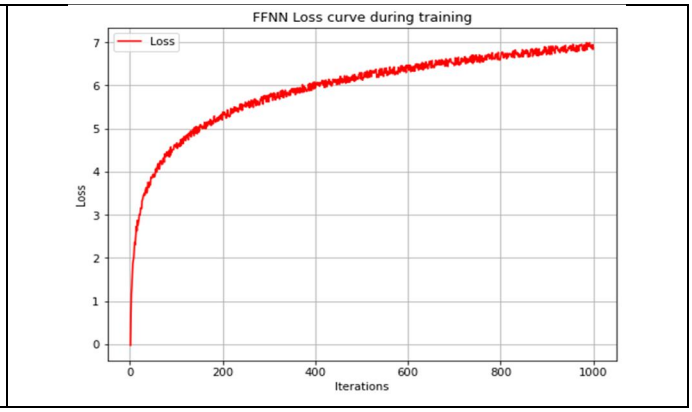


Fig6. FFNN Loss curve during training

Among the models tested, FFNN attained the highest accuracy but required significant computational resources, while SVM provided a balanced solution suitable for real-time applications.

B. Real-Time streaming performance

The system's real-time streaming feature was evaluated based on latency and throughput using Apache Kafka and Spark. These metrics are provided in Table 2 below. Thus, Kafka maintained low interference, with an average of 1,200 ms of delay in smaller batches and a maximum of 2,500 ms in larger batches. Micro batch data processing was also managed well under Spark, with an average time of 2500 ms for small micro batches and a maximum of 3500 ms for larger ones. The merging of Kafka and Spark showed promising results in terms of scalability and efficacy, as requested by the real-time application of rice quality prediction using precision agriculture

TABLE II. KAFKA-SPARK STREAMING LATENCY AND THROUGHPUT

Metric	Kafka	Spark Streaming
Average latency (small batch, ms)	1200 ms	2500 ms
Peak latency (large batch, ms)	2500 ms	3500 ms
Throughput (message/sec)	5000 messages/sec	3000 messages/sec
Fault tolerance high	High (automatic replication)	High (data persistence with RDDs)
Scalability	Horizontal scaling with partitions	Distributes nodes

C. Comparison with existing approaches

The fact that the proposed system will stream and incorporate scalability in real-time makes it a unique solution from the traditional batch processing techniques. Related works reviews on batch processing techniques have been addressed by previous researchers like Su Ying-xue [17] that did not involve real-time data streaming. They are mainly unsuitable for real-time operation, essential for precise farming. The Onesait platform was used to store and update data for real-time visualizations of the crops' health and quality predictions displayed on the operations dashboard.

V. DISCUSSION AND CONCLUSION

In this work, the ML models for real-time rice quality prediction were examined, and the best models shown in the table were FFNN, which had the added disadvantage of very high computational time. The problem with such a formulation is that it diminishes the realism of developing an FFNN in a real-life release, as timeframe and resources are critical. SVM provided a comparable balance between accuracy and computational costs, corroborating other studies stating that SVM is accurate in reproducibly moderate resource-constrained contexts, placing it well for precision agriculture solutions [18], [19]. On the other hand, we observed that K-Nearest Neighbours (KNN) tested moderate performance, and this is in agreement with studies that have pointed out that high-dimensional data has one major drawback: it is based on the classification by proximity. This supports that more complex algorithms are required when large-scale agricultural data are complex, as presented in [20]. They are switching to real-time data handling through Apache Kafka, and Spark aided in the processing, providing an edge over batch processing. Like earlier work, the present study established that big data technology can potentially enhance efficient and sustainable crop decision systems. While FFNN has the possible influence mentioned above, the model's computational burden makes the model unsuitable for large-scale applications. Further studies should look at building models somewhere in between, or at least less complex, resource-demanding architectures to achieve comparable levels of accuracy. However, extending the number of predictors added to the model, such as environmental conditions like moisture, temperature, and humidity, can also increase the accuracy and the system's applicability toward precision farming.

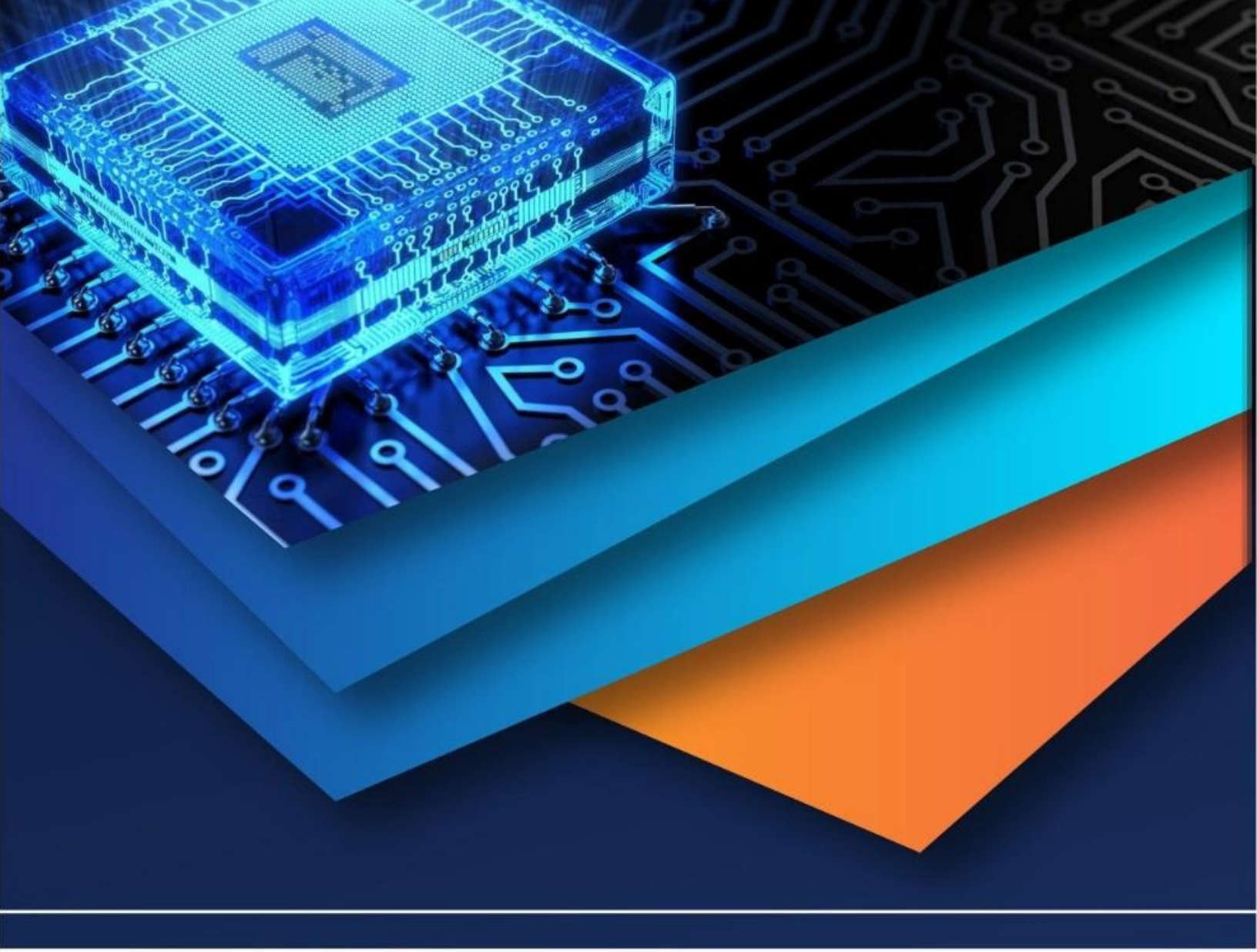
Last, we highlight how intelligent farming is made possible by machine learning and real-time data streaming; even now, there are issues with data processing and scalability. Cloud computing is characteristic of offering customizable solutions, but many systems do not fully leverage these services. This system advances the usefulness of rice quality prediction by integrating ML models with real-time streaming and cloud data management processes. This framework is generalizable well to other extensions of the administrative properties of agriculture and will ensure the effective utilization of resources.

REFERENCES

[1] Food and Agriculture Organizatio (FAO), “The future of food and agriculture: Trends and Challenges,” 2017.
[2] Md. Rayhan Shaheb, A. Sarker, and S. A. Shearer, “Precision Agriculture for Sustainable Soil and Crop Management,” in Soil Science - Emerging Technologies, Global Perspectives and Applications, 2022. doi: 10.5772/intechopen.101759.
[3] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, “Big Data in Smart Farming – A review,” 2017. doi: 10.1016/j.agsy.2017.01.023.
[4] J. W. Jones et al., “Brief history of agricultural systems modeling,” Agric Syst, 2017, doi: 10.1016/j.agsy.2016.05.014.
[5] T. Sofi, S. Mir, T. Mubarak, and Z. Ahmad Bhat, “Decision support systems in a global agricultural perspective-a comprehensive review. International Journal of Agriculture Sciences,” vol. 7, no. 1, pp. 403–415, 2015, [Online]. Available: http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217
[6] E. Elbasi et al., “Crop Prediction Model Using Machine Learning Algorithms,” Applied Sciences (Switzerland), 2023, doi: 10.3390/app13169288.
[7] F. Tian and Y. Zhang, “Spatiotemporal patterns of evapotranspiration, gross primary productivity, and water use efficiency of cropland in agroecosystems and their relation to the water-saving project in the Shiyang River Basin of Northwestern China,” Comput Electron Agric, vol. 172, May 2020, doi: 10.1016/j.compag.2020.105379.
[8] A. Cravero, S. Pardo, S. Sepúlveda, and L. Muñoz, “Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review,” 2022. doi: 10.3390/agronomy12030748.
[9] I. Attri, L. K. Awasthi, and T. P. Sharma, “Machine learning in agriculture: a review of crop management applications,” Multimed Tools Appl, vol. № 5, pp. 12875–12915, 2023, doi: 10.1007/s11042-023-16105-2.
[10] Mayur Rajaram Salokhe, “Machine Learning: Applications in Agriculture (Crop Yield Prediction, Diease and Pest Detection),” International Journal of Advanced Research in Science, Communication and Technology, pp. 592–597, Jul. 2023, doi: 10.48175/ijarsct-12088.
[11] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, “Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines,” Ore Geol Rev, 2015, doi: 10.1016/j.oregeorev.2015.01.001.
[12] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine Learning in Agriculture: A Review,” Sensors, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/s18082674.
[13] Z. H. Kok, A. R. Mohamed Shariff, M. S. M. Alfatni, and S. Khairunniza-Bejo, “Support Vector Machine in Precision Agriculture: A review,” 2021. doi: 10.1016/j.compag.2021.106546.
[14] Nodar Momtselidze, “Apache Kafka-Real-time Data Processing.”

[15] E. Giusti and S. Marsili-Libelli, "A Fuzzy Decision Support System for irrigation and water conservation in agriculture," Environmental Modelling and Software, 2015, doi: 10.1016/j.envsoft.2014.09.020.
[16] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big Data in Smart Farming – A review," 2017. doi: 10.1016/j.agsy.2017.01.023.
[17] Y. xue Su, H. Xu, and L. jiao Yan, "Support vector machine-based open crop model (SBOCM): Case of rice production in China," Saudi J Biol Sci, vol. 24, no. 3, pp. 537–547, Mar. 2017, doi: 10.1016/j.sjbs.2017.01.024.
[18] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," 2018. doi: 10.1016/j.compag.2018.02.016.
[19] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," 2018. doi: 10.1016/j.compag.2018.05.012.
[20] M. Bhavana and K. S. Rao, "Machine and Deep Learning-based Techniques for Precision Agriculture with Comparative Analysis," in 2nd International Conference on Automation, Computing and Renewable Systems, ICACRS 2023 - Proceedings, 2023. doi: 10.1109/ICACRS58579.2023.10404152.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)