Assessing Agricultural Land and Determining Crop Suitability for Optimum Yield using Machine Learning Approach

KOIRANGA, Abdullahi Hammanadama
Computer Science Department
Federal Polytechnic, Mubi, Adamawa State, Nigeria
koiranga.ab@gmail.com

Dr. Yusuf Musa Malgwi
Computer Science Department
Modibbo Adama University, Yola, Adamawa State, Nigeria
yumalgwi@mau.edu.ng

HAMIDU, Mohammed
ICT unit
Federal Polytechnic, Bali, Taraba State, Nigeria
mohammedhamidu1987@gmail.com
DOI: 10.56201/ijcsmt.v10.no4.2024.pg134.144

Abstract
Machine Learning has developed rapidly and increasingly becoming significant in agricultural domain as there is need of efficient logical practice for detecting new and valuable information in agricultural domain. Agriculture serves as the primary economic driver and major employment provider for the majority of the population in Nigeria. However, the sector is vulnerable to challenges stemming from inadequate agronomic practices. This study built an Agricultural land assessment model that predict optimum yield using logistic regression model and Support Vector Machine (SVM) classifiers using Python and VScode. Comparing Logistic Regression and Support Vector Machine (SVM) for classifying crops based on their nutrient content and yield, both algorithms exhibit high overall accuracy of around 87%. However, SVM stands out due to its consistently high precision, recall, and F1-scores across various crops. This indicates that SVM provides a more balanced and robust performance, vital for accurate agricultural predictions. The study also introduces a novel application of machine learning techniques to the field of agriculture and bridges the gap between traditional agricultural practices and modern data analytics, fostering innovation and improving agricultural efficiency and sustainability. The study can further enhance its impact on agricultural productivity and sustainability by Incorporate Internet of Things (IoT) devices to collect real-time data on soil moisture, temperature, and other environmental factors, promote improving the accuracy of crop suitability predictions. Also to extend the model to include a wider variety of crops and explore its applicability to different climatic and geographical regions, ensuring broader relevance and utility

Keywords*: Assessing, Agricultural Land, Crop Suitability, Optimum Yield, Machine Learning Approach*

Introduction

Farming, commonly referred to as agriculture, involves the cultivation of crops and the raising of animals. It plays a significant role in contributing to a country's economy by producing raw materials and food products. Industries utilize raw materials like cotton, produced through agriculture to manufacture various products used in everyday life. Agriculture serves not only as a means of food production but also as a source of resources for the creation of profitable products. The dominant method in the country is traditional farming, which relies on techniques passed down by experienced farmers. While these traditional techniques lack precision, they require considerable manual labour and time investment. (Durai & Shamili, 2022).

Agriculture serves as the primary economic driver and major employment provider for the majority of the population in Nigeria. However, the sector is vulnerable to challenges stemming from inadequate agronomic practices, a lack of market connectivity, and a scarcity of skilled professionals offering guidance to farmers (Ibiyemi, 2022).

Farmers frequently rely on agricultural specialists and advisors for information regarding the selection of farmland. Regrettably, these agricultural experts are not consistently accessible when farmers require guidance on choosing suitable farmland (Ibiyemi, 2022).

The process of farm selection involves choosing appropriate crops for a farm based on physical, economic, and climatic factors related to the specific farm unit. These factors significantly impact crop yields, presenting challenges that farmers encounter during the planning of their cropping strategies (Liliane & Charles, 2020).

Information can be obtained through various means, including conducting interviews with domain experts, analysing documents, and making observations. However, due to the personal nature of tacit knowledge and the possibility that experts may not fully articulate all their insights during interviews, there remains undisclosed information about the problem (Liliane & Charles, 2020).

To address this challenge, the suggestion is to employ the Crop Selection Method Using Machine Learning Techniques, which can effectively tackle the shortage of experts and the limited spread of information. An expert system, a potent tool in agriculture, is proposed for this purpose. An expert system is software designed to emulate the problem-solving abilities of one or more human experts within a specific domain, falling under the umbrella of artificial intelligence. Also known as a Knowledge-Based System (KBS), an Expert System (ES) is a computer program made to replicate the decision-making behaviour of an expert in a particular domain or discipline. Such an expert system could be developed to facilitate decision-making and the dissemination of technology tailored to specific locations. (Aguboshim & Otuu, 2023).

Statement of the Problem

Conventional approaches in evaluating agricultural land and identifying suitable crops typically depend on limited data and subjective assessments, often resulting in less ideal crop yields and inefficient land utilization. There is a demand for a more accurate and data-driven method to assess land characteristics and determine the most appropriate crops for particular areas. This research intends to apply machine learning techniques to examine factors like soil properties, climate conditions, and historical crop data to create a model that can precisely predict crop suitability and maximize yield, thereby improving agricultural productivity and sustainability

Aim and Objectives of the Study

The aim of this study is to leverage machine learning to optimize agricultural practices by aiding informed decision making regarding crop selection and land use, thereby maximizing yield and sustainability. The objectives of the study include:

i. Gather diverse dataset encompassing soil nutrient and crop yields.
ii. Identifying and extracting essential features influencing crop growth and land suitability for best agricultural practices
iii. Employ machine learning algorithms Logistic Regression and Support vector Machine (SVM) to build predictive models for assessing land suitability for various crops based on gathered data
iv. To evaluate the performance of the two techniques in order to select to best

Review of Related Literature

Jadhav & Bhaladhare, (2022) uses support vector machine (SVM), artificial neural network (ANN), random forest (RF), multivariate linear network (MLN), and a combination of regression and KNN for a system for crop recommendation. The study suggest that support vector machine is best approach for its minimal computational endeavours despite the lack of geospatial analysis that incorporates and utilises all available seasonal, soil, weather, and temperature data for effectiveness and accuracy.

Jha et al (2020) published on a machine learning approach to recommend suitable crops and fertilizers for agriculture which uses case-based approach to separate various land properties for agriculture. The study was limited to selection of crops on soil texture.

Akulwar, (2020) is of opinion that determines optimum cultivation date and evaluated demand & supply risk are best factors to be considered for crop selection. Using random forest algorithm, the study lacks concentration crop diseases and pest control measures.

Sharma et al., (2021) designed artificial intelligent farm using decision tree, gaussian naive bayes, logistic regression, random forests and xgboost to explore soil nutrients that may induce learning productivity deployed on android platform which makes the system platform dependent.

Priya & Yuvaraj, (2019) uses deep learning to create internet of a thing based gradient descent approach for precision crop suggestion using MLP. The system hardware was successfully developed. The study further proposes the use of hybrid approaches for suggesting fertilizers to be added in a timely manner for achieving high profit and yield.

Cockburn, (2020) reviewed on application and prospective discussion of machine learning for the management of dairy farms. A review that thoroughly examine over 90 papers between 2015 and 2020 reveals machine learning algorithms have become common tools in most areas of dairy research, particularly to predict data. Despite the quantity of research available, most tested algorithms have not performed sufficiently for a reliable implementation in practice.

(Elbasi et al., 2023) worked on crop prediction model using machine learning algorithms precisely support vector machine the system however, is limited to publish dataset which may enquire adequate use of hardware gadgets such as GPS-based IoT and sensor data from different geographic regions for implementation.

Anguraj et al., (2021) publishes a paper on crop recommendation on analysing soil using soil moisture, temperature, humidity and pH are the model's input parameters to decision support system of IoT. The model however does not surpass the theoretical phase.

Choudhary et al., (2022) contributed to the field by writing on crop recommendation system and plant disease classification using machine learning for precision agriculture using heap map matrix exploration.

Gosai et al., (2021) uses decision tree (DT), naïve bayes (NB), support vector machine (SVM), logistic regression (LR), and random forest (RF) machine learning algorithms to create a crop recommendation system. The proposed model is pending implementation.

Rakhra et al., (2022) proposed a system model using machine learning for smart farming to forecast farmers' interest in hiring equipment using logistic regression k-neighbours classifier and decision tree machine learning algorithm the system suggests the exploitation of sensors. Another contribution by (Priyadharshini et al., 2021) proposes a machine learning model on intelligent crop recommendation system using reinforcement learning algorithm. Similarly, (Sharma et al., 2020) embark on developing a LM model using reinforcement learning for precision agriculture.

Vincent et al., (2019) invented alternative strategy for accumulating agricultural farm information by developing sensors driven ai-based agriculture recommendation model for assessing land suitability using reinforced machine learning algorithm. Motwani et al., (2022) like other literatures; the study dwells on soil analysis and crop recommendation

Conceptual Framework

The conceptual framework described the visual representation in form of a diagram that illustrates the interrelationships between key concepts and variables. It provides a structured framework for research that organizes the thoughts and guide analysis. The framework and methodology of the proposed model are presented in Figure 1.

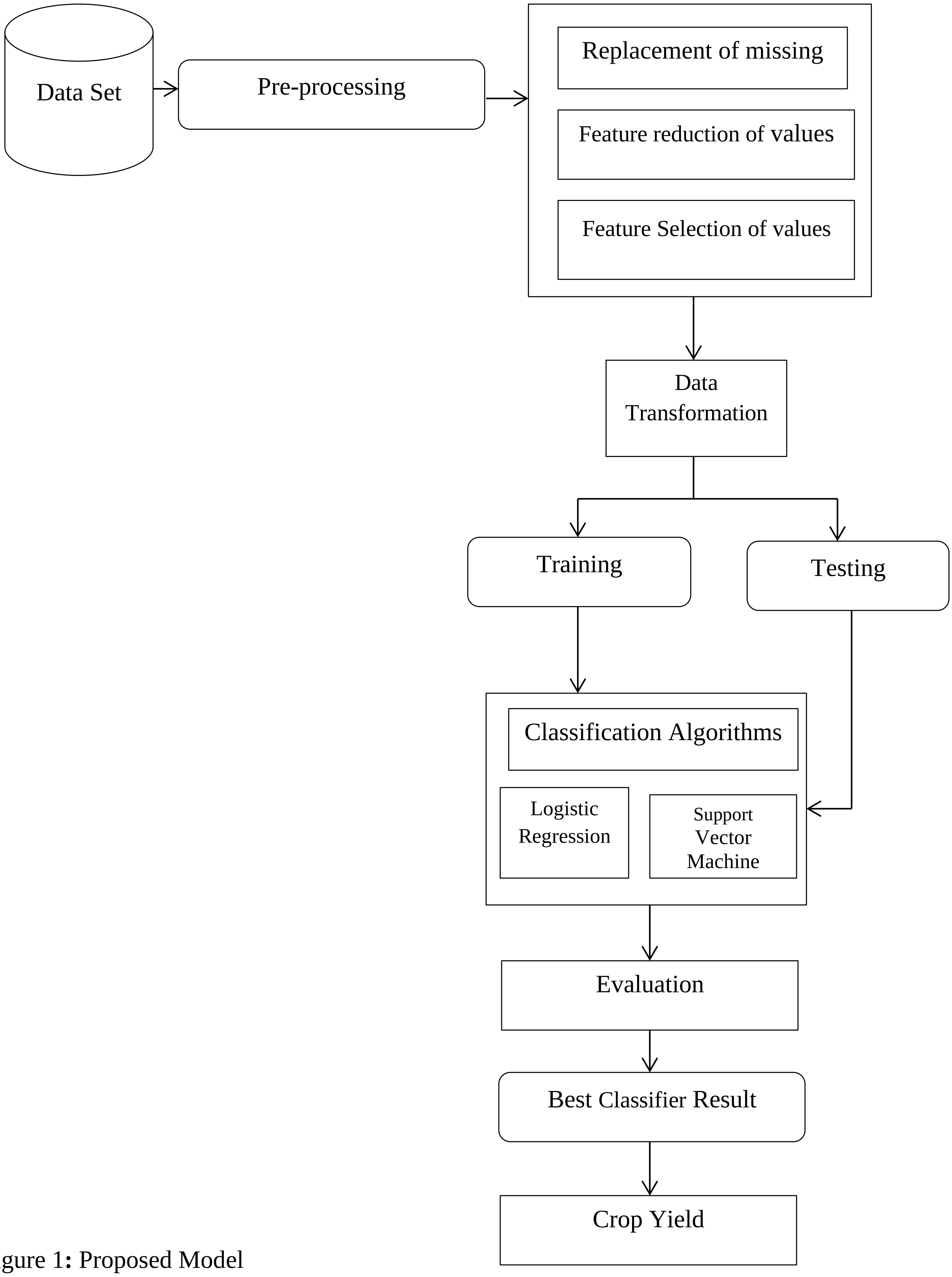

Figure 1: Proposed Model

1. Dataset

The dataset consists of parameters like Nitrogen(N), Phosphorous(P), Pottasium(K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Manganese (Mn) and Iron (Fe). The data set has 1112 instance or data that have taken from the www.hub.arcgis.com for building and testing the model. The dataset include sixteen different crops such as Bambara groundnut, Cassava, Cowpeas, Groudnut, Maize, Millet, Okra, Onion, Peanut, Paper, Sorghum, Rice, Sugarcane, Sweet potatoes, Tomatoes and Yam

2. Data Pre-processing

Data pre-processing refers to the steps and techniques applied to raw data before it can be used for analysis or modelling. It is an essential part of the data preparation stage in data science and machine learning workflows. The goal of data pre-processing is to clean, transform, and enhance the raw data to improve its quality, consistency, and convert the data into a suitable form that can be used by algorithms.

For the successful application pre-processing the data which is acquired is in raw form. It contains some incomplete, redundant, inconsistent data. The redundant data was filtered and normalized. The pre-processed data was stored in a 'Comma Separated Values' (CSV) file format. Three main pre-processing steps have been applied to the dataset which are data cleaning, feature encoding, and feature selection

3. Feature Extraction

This step is focus on identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of classifiers

4. Analysis Setup

The Soil nutrient data were collected from the www.hub.arcgis.com containing about 1112 soil records consists of parameters like Nitrogen(N), Phosphorous(P), Pottasium(K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Manganese (Mn) and Iron (Fe) with their respective quantity and their corresponding appropriate crops yields as shown in figure 4.1.

	N (lbs/acre)	P (lbs/acre)	K (lbs/acre)	Ca (ppm)	Mg (ppm)	Cu (ppm)	Mn (ppm)	Fe (ppm)	Crop and Yield
0	15.0	3.0	4.0	2.0	10.0	0.010	0.05	0.20	Rice
1	12.0	4.0	8.0	3.0	15.0	0.020	0.08	0.30	Tomatoes
2	8.0	2.0	6.0	1.5	12.0	0.010	0.06	0.25	Yams
3	10.0	3.0	5.0	2.0	8.0	0.015	0.04	0.18	Okra
4	7.0	1.5	4.0	1.0	6.0	0.008	0.03	0.15	Cassava

Figure 2: Dataset description

5. Building the Model

The model included several stages and was built by importing necessary libraries and loaded the dataset. The features of the target variable were separated and features were standardized. Logistic regression and Support Vector Machine models were trained and evaluated its performance by checking accuracy and creating a classification report. Lastly, the main function was set up to load the data, pre-process it, train the model, evaluate it, and save the results, ensuring everything runs smoothly when the script is executed.

6. Model Training and Testing

In order to train the classifiers models, 80% of Dataset with known output (label) were used for training and the other 20% unlabelled data for testing. Using logistic regression and support vector machine models, figure 3 and figure 5 shows the results

Results and Discussions

1. Using Logistic regression on testing dataset

The model produced by logistic regression testing set is shown in figure 3.

Model accuracy: 0.8579136690647482
Classification Report:

	precision	recall	f1-score	support
Bambara Groundnut	0.93	0.82	0.88	68
Cassava	0.80	0.86	0.83	71
Cowpeas	0.83	0.85	0.84	65
Groundnut	0.84	0.75	0.80	65
Maize	0.82	0.77	0.79	65
Millet	0.82	0.93	0.87	68
Okra	0.88	0.83	0.86	71
Onions	0.89	0.89	0.89	72
Peanuts	0.78	0.87	0.83	71
Peppers	0.86	0.90	0.88	68
Rice	0.75	0.81	0.78	75
Sorghum	0.81	0.88	0.84	66
Sugarcane	1.00	0.93	0.96	71
Sweet Potatoes	0.90	0.91	0.91	68
Tomatoes	0.94	0.95	0.94	76
Yams	0.92	0.76	0.83	72

Figure 3: Logistic Regression Classification Report.

The performance of a machine learning model in predicting crop suitability for various crops

i. The precision for **Sugarcane** is 1.00 that is means all predictions for Sugarcane was correct. However, for **Peanuts**, the precision is 0.78, indicating some false positives (other classes being predicted as Peanuts).

ii. **Recall**: High recall for a class means the model successfully identifies most instances of that class. For instance, **Tomatoes** have a recall of 0.95 that means the model correctly identifies 95% of all Tomato instances. In compare, **Yams** have a recall of 0.76, indicating that 24% of the Yams instances were not identified correctly.

iii. **Millet** has a high f1-score of 0.87, suggesting good performance, while **Groundnut** has a lower f1-score of 0.80, pointing to potential issues in balancing precision and recall.

iv. **Support**: This indicates the number of true instances for each crop in the dataset. The support values range from 65 to 76, indicating a relatively balanced dataset across different crops.

2. The performance Evaluation

accuracy			0.86	1112
macro avg	0.86	0.86	0.86	1112
weighted avg	0.86	0.86	0.86	1112

C:\Users\Muhammad\Desktop\Projects 2024\koiranga\software>

Figure 4: Logistic Regression Performance evaluation

Accuracy: The accuracy of the model is 0.86, means the model correctly predicts the crop suitability 86% times. This is a solid performance, indicating that the model is generally reliable.
Macro Average: Both the macro average precision and recall are 0.86. This suggests that, on average, the model performs well across all crops without being biased towards any particular crop.
Weighted Average: The weighted average precision and recall are also 0.86. The weighted average takes into account the support of each class, giving more weight to the classes with more instances. This reinforces that the model maintains consistent performance even when considering the distribution of the crops in the dataset.

3. Using Support Vector Machine on testing dataset

The model produced by Support Vector Machine (SVM) testing set is shown in figure 5

Model accuracy: 0.8758992805755396
Classification Report:

	precision	recall	f1-score	support
Bambara Groundnut	0.76	0.91	0.83	68
Cassava	0.91	0.82	0.86	71
Cowpeas	0.94	0.92	0.93	65
Groundnut	0.93	0.86	0.90	65
Maize	0.86	0.88	0.87	65
Millet	0.83	0.99	0.90	68
Okra	0.84	0.82	0.83	71
Onions	0.93	0.94	0.94	72
Peanuts	0.80	0.80	0.80	71
Peppers	0.85	0.99	0.91	68
Rice	0.91	0.80	0.85	75
Sorghum	0.88	0.85	0.86	66
Sugarcane	1.00	0.83	0.91	71
Sweet Potatoes	0.89	0.91	0.90	68
Tomatoes	0.87	0.95	0.91	76
Yams	0.90	0.76	0.83	72

Figure 5: Support Vector Machine Classification Report

4. The performance Evaluation

accuracy			0.88	1112
macro avg	0.88	0.88	0.88	1112
weighted avg	0.88	0.88	0.88	1112

Figure 6: Support Vector Machine Performance evaluation

The model has an accuracy of 0.88 or 88%, This means that the model correctly predicts the class of the crops in 88% of the instances in the test dataset.

5. Models Comparison Experimental Results

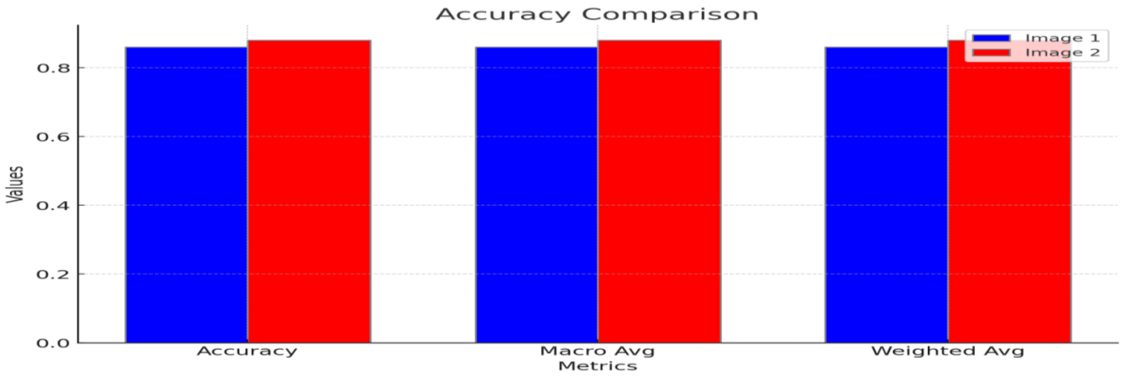

Figure 7**:** Model Comparison

The blue bars represent the values from the logistic regression model, and the red bars represent the values from the second image. This shows that accuracy improved from 0.86 to 0.88 and both macro and weighted averages also increased from **0.86** to **0.88**.

Table 1: Result Analysis

Algorithm	Accuracy
Logistic Regression (LR)”	**86%**
“Support Vector Machine (SVM)”	**88%**

Findings

Comparing Logistic Regression and Support Vector Machine (SVM) for classifying crops based on their nutrient content and yield, both algorithms shows high accuracy of around 87%. However, SVM stands out due to its consistently high precision, recall, and F1-scores across various crops. This indicates that SVM provides a more balanced and robust performance for accurate agricultural predictions.

SVM is well-matched for handling the intricate relationships between soil nutrients and crop yields. The kernel trick enables SVM to model non-linear relationships effectively, making it a preferable choice for datasets with numerous features and complex interactions.

Conclusion

The analysis shows that both Logistic Regression and SVM achieve similar accuracy SVM offers superior performance in terms of precision and recall across most crops. SVM's ability to balance these metrics makes it more effective for practical applications in agricultural predictions. The algorithm's robustness against over fitting and its capacity to model non-linear relationships between soil nutrients and crop yields further enhance its suitability for this task. Therefore, SVM is recommended as the preferred algorithm for classifying crop nutrients and their yielding capacity.

Recommendations

i. The system provides clear recommendations on crop suitability to farmers and agricultural planners

ii. It is recommended to use the Support Vector Machine algorithm for classifying crop nutrients and predicting yields due to its balanced and robust performance.

iii. SVM is recommended, exploring other advanced algorithms such as ensemble methods or deep learning could provide additional improvements in classification accuracy and predictive power.

iv. Continuously gathering more data will improve the model's learning and prediction capabilities, which ensures up-to-date with changing agricultural conditions.

References

Aguboshim, F. C., & Otuu, O. O. (2023). Using computer expert system to solve complications primarily due to low and excessive birth weights at delivery: Strategies to reviving the ageing and diminishing population. *World Journal of Advanced Research and Reviews*, *17*(3), 396–405.

Akulwar, P. (2020). A Recommended System for Crop Disease Detection and Yield Prediction Using Machine Learning Approach. In S. N. Mohanty, J. M. Chatterjee, S. Jain, A. A. Elngar, & P. Gupta (Eds.), *Recommender System with Machine Learning and Artificial Intelligence* (1st ed., pp. 141–163). Wiley. https://doi.org/10.1002/9781119711582.ch8.

Anguraj, K., Thiyaneswaran, B., Megashree, G., Shri, J. P., Navya, S., & Jayanthi, J. (2021). Crop recommendation on analyzing soil using machine learning. *Turkish Journal of Computer and Mathematics Education*, *12*(6), 1784–1791.

Choudhary, M., Sartandel, R., Arun, A., Ladge, L., Hiranwal, S., & Mathur, G. (2022). Crop Recommendation System and Plant Disease Classification using Machine Learning for Precision Agriculture. *Artificial Intelligence and Communication Technologies, SCRS*, 39–49.

https://www.publications.scrs.in/uploads/final_menuscript/7e6d4680313ad60e32d6e3f82d98 1a8e.pdf

Cockburn, M. (2020). Application and prospective discussion of machine learning for the management of dairy farms. *Animals*, *10*(9), 1690.

Cropland Area by Country—Worldometer. (n.d.). Retrieved December 25, 2023, from https://www.worldometers.info/food-agriculture/cropland-by-country/

Dandotiya, B., & Sharma, H. K. (2022). Climate change and its impact on terrestrial ecosystems. In *Research Anthology on Environmental and Societal Impacts of Climate Change* (pp. 88–101). IGI Global.

https://www.igi-global.com/chapter/climate-change-and-its-impact-on-terrestrial-ecosystems/293895

Durai, S. K. S., & Shamili, M. D. (2022). Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, *3*, 100041.

Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, *13*(16), 9288.

https://pdfs.semanticscholar.org/6e08/7108aa8048da8cfc82cdecb7071a55bab488.pdf **Esri.** (n.d.). *ArcGIS Hub*. Retrieved July 6, 2024, from https://hub.arcgis.com/

Gosai, D., Raval, C., Nayak, R., Jayswal, H., & Patel, A. (2021). Crop recommendation system using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *7*(3), 558–569.

Ibiyemi, M. A. (2022). *Enhancing the Productivity of the Nigerian Agricultural Sector Through Innovation and Technology* [PhD Thesis, Northcentral University]. https://search.proquest.com/openview/f0bcd389eae7c3e1a519b9eb1104e92c/1?pq-origsite=gscholar&cbl=18750&diss=y

Jadhav, R., & Bhaladhare, P. (2022). A Machine Learning Based Crop Recommendation System: A Survey. *Journal of Algebraic Statistics*, *13*(1), 426–430.

Jha, G. K., Ranjan, P., & Gaur, M. (2020). A Machine Learning Approach to Recommend Suitable Crops and Fertilizers for Agriculture. In S. N. Mohanty, J. M. Chatterjee, S. Jain, A. A. Elngar, & P. Gupta (Eds.), *Recommender System with Machine Learning and Artificial Intelligence* (1st ed., pp. 89–99). Wiley. https://doi.org/10.1002/9781119711582.ch5

Liliane, T. N., & Charles, M. S. (2020). Factors affecting yield of crops. *Agronomy Climate Change & Food Security*, 9.

Priya, P. K., & Yuvaraj, N. (2019). An IoT based gradient descent approach for precision crop suggestion using MLP. *Journal of Physics: Conference Series*, *1362*(1), 012038. https://iopscience.iop.org/article/10.1088/1742-6596/1362/1/012038/meta

Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 843–848. https://ieeexplore.ieee.org/abstract/document/9418375/

Rakhra, M., Sanober, S., Quadri, N. N., Verma, N., Ray, S., & Asenso, E. (2022). Implementing machine learning for smart farming to forecast farmers' interest in hiring equipment. *Journal of Food Quality*, *2022*. https://www.hindawi.com/journals/jfq/2022/4721547/

Sharma, A., Bhargava, M., & Khanna, A. V. (2021). AI-Farm: A crop recommendation system. *2021 International Conference on Advances in Computing and Communications (ICACC)*, 1–7. https://ieeexplore.ieee.org/abstract/document/9708104/

Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, *9*, 4843–4873.

Vincent, D. R., Deepa, N., Elavarasan, D., Srinivasan, K., Chauhdary, S. H., & Iwendi, C. (2019). Sensors driven AI-based agriculture recommendation model for assessing land suitability. *Sensors*, *19*(17), 3667.