International Journal of Advances in Applied Sciences (IJAAS)
Vol. 14, No. 4, December 2025, pp. 999~1009
ISSN: 2252-8814, DOI: 10.11591/ijaas.v14.i4.pp999-1009

Harvesting insights: exploring machine learning for crop selection and predictive farming

Tanvi Deshmukh[1], Anand Singh Rajawat[1], Amol Potgantwar[2]
[1]Department of Computer Science and Engineering, School of Computer Science and Engineering, Sandip University, Nashik, India
[2]Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Sandip Foundation, Nashik, India

Article Info

Article history:

Received Oct 14, 2024
Revised Jul 28, 2025
Accepted Sep 21, 2025

Keywords:

Crop selection
Machine learning
Modern agriculture
Optimization
Predictive farming

ABSTRACT

Modern agriculture has undergone a significant evolution, adopting advanced techniques to streamline crop management processes. One such advancement is the integration of machine learning (ML) technology, which shows great promise in optimizing crop selection and enhancing economic returns. Key determinants of crop productivity, including water availability, soil quality, weather conditions, and timely resource allocation, play pivotal roles in the farming ecosystem. Harnessing these factors, ML algorithms facilitate the identification of optimal crop choices and provide continuous monitoring of cultivation processes to anticipate evolving crop needs. This paper investigates various ML techniques employed for crop selection and evaluates their effectiveness in agricultural settings. Through a comparative analysis, we highlight the advantages of these techniques and provide insights into their potential impact on current farming management practices. By leveraging ML for predictive farming, stakeholders can make informed decisions to maximize yields, minimize resource wastage, and promote sustainable agricultural practices. This study contributes to the ongoing discourse on the integration of technology in agriculture and underscores the transformative potential of ML in shaping the future of crop management. We investigate recent papers from the years 2020 to 2024.

Corresponding Author:

Tanvi Deshmukh
Department of Computer Science and Engineering, School of Computer Science and Engineering
Sandip University
Nashik 422213, Maharashtra, India
Email: tanvid532@gmail.com

1. INTRODUCTION

Modern technology is crucial to fulfilling several day-to-day life needs. As the world has a high priority to increase food production for a large population, crop management is the way to be used efficiently. At this point, human decision-making limitations can be overcome with the expert support of artificial intelligence. Several aspects of crop growth integrate a proper crop selection to healthy growth that depends on soil quality, temperature conditions, weather conditions, and water availability based on site. Thus, specific information about a particular location can help to schedule a suitable crop with the help of the whole analysis. Machine learning (ML) techniques are trained to analyze such data, and prediction is obtained according to it. The software carries all processes and acts like an expert in decision-making. Therefore, modern farming is introduced to ML techniques for higher productivity and enhanced economic benefits [1].

In this review paper, we investigate how ML is influencing the agricultural sector, with an emphasis on predictive farming and crop selection. The accuracy and flexibility needed to adjust to changing environmental circumstances are typically lacking in traditional methods of crop management, which depend on manual analysis and historical data. On the other hand, ML approaches give a data-driven model for selecting crops, taking several factors such as soil properties, climate conditions, and availability of resources into account.

In this paper, the efficiency of different ML methods for agricultural applications is studied through an extensive literature survey and empirical study. We strive to discover the optimal manner of implementing ML to farming systems by counterposing strategies to each other, so that the advantages and disadvantages of each strategy become obvious to us. The paper highlights the potential for ML approaches to enhance the identifiability of important characteristics of agricultural management, such as crop selection, crop yield predictions, and disease detection. These methods have the advantage of being able to process very large datasets and to find patterns that are not detected by standard approaches. It understands the critical enablers of crop performance, such as the available moisture, condition of the soil, the conditions of the monsoons, and the when-then-not application of inputs. These features can easily be analyzed by ML and can result in possible efficient crop selection.

The study provides a comparative discussion on all the considered ML methods in terms of their merits and demerits as applied to agricultural data. Such an analysis can be useful to figure out what approaches are most useful for which operations within agriculture. The study underscores the importance of implementing ML technologies within agricultural ecosystems, recommending that the precision agriculture lobby accessory should include tools to optimize the best input application and efficiency through sustainability.

The organization of this paper is as follows: section 2 presents the basics of ML and its applicability for agriculture. Section 3 explains the determinants of crop choice and the application of ML in making crop selection more efficient. In section 4, we compare contextual, aggregate, statistical, and ML methods for predictive farming, detailing the pros and cons of individual models. Section 5 describes the common ML methods applied in agriculture and their pros and cons. Practical challenges and opportunities in the integration of ML into agricultural practices are discussed in section 6. The paper then concludes with a summary of the main findings, implications for future research, and recommendations for agricultural stakeholders in section 6.

2. RELEVANCE OF ML TO AGRICULTURE

A subfield of artificial intelligence, it entails training computers to learn on their own, without being explicitly programmed. ML has been increasingly attracting attention from the agricultural communities in the hope of improving efficiency, sustainability, and production in a variety of tasks. A short overview of ML and its uses in agriculture is presented by [1], [2]:

- Data-driven approach: ML systems crunch enormous amounts of data in search of patterns that will produce predictions. Any different data could have also been used to train the ML models in agriculture. These sources can be weather stations, soil sensors, satellite images, or farm equipment.
- Supervised learning: supervised learning enables the training of models with labelled data by relating input features (e.g., weather data, soil properties) to corresponding output labels (e.g., crop yield, disease). Tasks such as weed identification, disease detection, and forecasting of agricultural production are performed using this technique.
- Unsupervised learning: here, one trains models on unlabelled data in order to potentially discover patterns or to group similar points together in unsupervised learning. Grouping similar crop fields together, identifying abnormal growth, and sectioning fields based on soil attributes are just a few agricultural uses for unsupervised applications.
- Reinforcement learning: one subset of ML, reinforcement learning, instructs agents to act in such a manner as to maximize some cumulative reward. Utilize reinforcement learning in farming to irrigate more optimally, work crops more effectively, and distribute resources more sensibly.
- Deep learning: deep learning is a ML subfield that attempts to learn hierarchical features from data using networks with multiple layers. Convolutional neural networks (CNNs) and other deep learning models for agriculture may be used for purposes like weed detection, yield prediction by employing remote sensing data, and crop disease diagnosis from images.
- Predictive analytics: ML can enable agricultural predictive analytics by sorting through past data and current environmental conditions, predicting such outcomes as crop yields, pest infestations, and market pricing. With these predictions, farmers can streamline resource allocation, schedule planting and harvest times, and more.
- Precision agriculture: precision agriculture, in which farm management procedures are customized to individual plants or tiny parts of fields, relies heavily on ML. ML algorithms maximize yields by examining weather, crop health data, and spatial and temporal variability in inputs such as water,

fertilizer, and pesticides. ML algorithms optimize yields while minimizing negative environmental impacts.
- Automation and robotics: ML algorithms have fed automated farming systems, robotic tools that plant, weed, and harvest crops. Such technological developments provide the ability to continuously monitor and apply to farming practices, and labor costs are minimized, and efficiency is enhanced.

ML is bringing about a new age of smart farming practices by facilitating data-driven decision-making, improving resource consumption, increasing productivity, and reducing environmental impact.

3. ROLE OF ML IN OPTIMIZING CROP SELECTION PROCESS

Crop selection is a critical decision for farmers and is influenced by various factors that impact the success and profitability of agricultural operations [1]–[4]. The key factors influencing crop selection include:
- Climate and weather conditions: different crops have specific requirements for temperature, precipitation, humidity, and sunlight. Farmers must consider the climate and weather patterns of their region to select crops that are well-suited to the prevailing conditions.
- Soil characteristics: soil fertility, texture, pH levels, drainage, and nutrient content play a crucial role in crop growth and yield. Farmers analyze soil test results to determine which crops will thrive in their soil conditions.
- Market demand and prices: market demand for certain crops and their selling prices influence farmers' decisions on what to plant. They consider factors such as consumer preferences, market trends, and potential profitability when selecting crops.
- Water availability and irrigation: crop selection is heavily influenced by water availability and irrigation infrastructure. Farmers choose crops that can be adequately irrigated within the constraints of water availability, whether from rainfall, surface water, or groundwater sources.
- Pest and disease resistance: some crops are more susceptible to pests and diseases than others. Farmers consider the prevalence of pests and diseases in their region and select crop varieties that have natural resistance or tolerance to common threats.
- Crop rotation and diversity: crop rotation is essential for maintaining soil health, managing pests and diseases, and improving overall crop yields. Farmers select crops based on their compatibility with rotation schedules and the benefits they provide to soil fertility and structure.
- Labor and equipment availability: the availability of labor and farm machinery influences crop selection decisions. Farmers choose crops that can be efficiently planted, managed, and harvested with the available resources, labor force, and equipment.
- Government policies and incentives: government policies, subsidies, and incentives can influence crop selection decisions. Farmers may choose crops eligible for support programs, such as crop insurance, price supports, or conservation incentives.

ML plays a crucial role in optimizing the crop selection process by leveraging data-driven approaches to analyze and model complex interactions between these factors. Here's how ML helps optimize crop selection:
- Data analysis and prediction: ML algorithms sift through mountains of data on weather, soil, crop yields, and market pricing to find trends and make predictions about which crops would thrive in certain environments.
- Recommendation systems: farmers can get tailored crop suggestions via ML-based recommendation systems that take into account their specific location, soil type, weather, and other variables. In order to maximize profits and yields, these systems assist farmers in making educated decisions.
- Risk assessment and management: ML models assess and quantify the risks associated with different crops, such as susceptibility to pests, diseases, or adverse weather events. Farmers can use this information to mitigate risks and optimize their crop selection strategies.
- Optimization algorithms: ML optimization algorithms help farmers maximize their returns by considering multiple objectives, such as yield, profit, resource use efficiency, and environmental sustainability. These algorithms recommend crop mixes and planting strategies that balance competing priorities.
- Real-time monitoring and decision support: AI allows for the tracking of weather, crop health, and market movements in real-time. In order to adapt their crop selection and management procedures to evolving conditions, farmers are provided with timely alerts, suggestions, and decision support.

All things considered, ML gives farmers more agency when it comes to data-driven decision-making, which in turn improves crop selection results and maximizes agricultural sustainability, resilience, and production.

4. LITERATURE REVIEW

Agriculture is complex to predict the appropriate crop for selection, and growth in agriculture features several aspects that impact the growth and production of plants. These parameters are heavily investigated, and various models are suggested to facilitate the task. These studies introduced new approaches based on temperature, water, soil, and weather information for estimating the best option for an improved production and growth management schedule. In this direction, the present study intends to revolutionize traditional farming practices by providing data-driven analysis and ML ML-based crop fertilizing system [5]. The approach increases agricultural sustainability and productivity by collecting various suggestions for crop selection and fertilization based on ML techniques. The problem of predicting old and future agricultural yields using ML techniques is also addressed in this paper [6]. For future agricultural planning and decision-making, ML models help in predicting crop yield accurately by analyzing environmental conditions and historical data.

The authors of his research [7] propose to use ML algorithms for building a crop prediction model. The model accurately predicts crop output for farmers and other agricultural stakeholders by integrating various environmental and agricultural factors. For making a decision on the crop or to predict the yield, an ML model is introduced in this work [8]. The proposed approach does farm management, optimizes cropping, and productivity by considering environmental conditions and historical data to support farm operators in crop selection and yield estimation. To model red crown rot (RCR) and soybean, remote sensing is combined with crop modeling based on ML methods in this work [9]. Crop modeling and forecasting are well supported by the method's embrace of numerous data sources and predictive algorithms with potential for better crop management decisions. This research [10] is about employing ML associated with remote sensing technologies in strawberry cropping phenotyping and management.

This study suggests that the application of ML algorithms and remote sensing information could enhance crop monitoring, disease detection, and yield estimation in strawberry production. It is also proposed a ML method [11] for the crop selection and yield estimation. The model that lets the farmers and anyone else working with agriculture make better decisions, as it takes into account a range of agricultural data and environmental conditions, which also helps in crop selection and yield prediction. A number of feature selection techniques and classifiers are employed in this work [12] for studying crop prediction in the context of the attributes of the agricultural background. The research aims to support agricultural planning and decision-making by accurately predicting crop yield from agricultural data and an ML algorithm.

For smart agriculture crop production prediction, a hybrid-Deep Learning system is presented in this paper [13]. The proposed model enables preemptive decision-making and resource allocation in agri-management by predicting crop yields with high accuracy by integrating deep learning approaches and agricultural data. For modelling and predicting maize crop yield, the application of artificial neural network methodology is studied in this work [14]. This work aims to contribute to better agricultural planning and decision-making by generating a robust model using a model of artificial neural network to estimate maize crop yield in a more precise way, with historical data and environmental variables being evaluated. Combining ML algorithms for selecting crops based on weather patterns is studied in this work [15]. The ultimate goal of the study is to improve agricultural productivity and management practices by the development of crop selection processes based on weather analysis assisted by environmental factors, with the use of meteorological data and ML algorithms.

The objective of the current study [16] is to use an artificial neural network approach for the prediction and forecasting of Pakistan's wheat production. In this research, an artificial neural network model is developed to predict wheat production using historical data and environmental factors. The model will enable wheat growers to make better decisions and use resources more efficiently. Applying a modified recursive feature elimination with several classifiers, this study [17] attempts to predict the suitability of soil for crop growth by using soil and environmental attributes.

The purpose of this study is to contribute to the optimization of agricultural land use and management by evaluating environmental and soil data to predict on what locations which crops could be grown. A framework to increase accuracy in crop production estimation is presented in this work [18] in the context of the internet of things (IoT) and ML-based precision agriculture, named PEnsemble 4. ML algorithms are combined with IoT data to enable better crop yield forecasts. This capability leads to superior decision-making in precision agriculture. Predicting agricultural crop yield using ML and deep learning techniques is presented in [19]. The main objectives of the study are to enhance the crop yield prediction for better agricultural planning and resource management through the analysis of agricultural data and the use of advanced ML techniques.

An agricultural crop yield prediction analysis based on ML techniques is also presented in this work [20]. The objective of this project is to improve crop yield estimation methods with the use of ML algorithms and to improve resource management. Forecasting agricultural yields with ML models is the topic of this work [21], and applies to maize and potatoes sown in Ireland. To aid farmers and other actors of the agricultural sector in making better judgments, the project will study agricultural data and employ ML algorithms in predicting crop yields. For the prediction of agricultural productions in Quezon Province, Philippines, the recent [22] compares various ML models.

The ultimate aim of the study is to enable the farmers to make better decisions, as the study determines the best predictive ML algorithm for predicting regional crop production. Hybrid deep learning models have been proposed in [23] to forecast agricultural production. The clipping is useful for future planning in an agriculture system supported by combining deep learning with the agricultural data, forming hybrid models that help to predict the crop's production with high efficacy. In Ethiopia's Lower Kulfo Watershed, in another research [24], deep learning algorithm practices, along with support vector machine (SVM), and to predict crop productivity have been investigated. The aim of the study is to enhance regional agricultural planning and decision-making by enhancing crop production Forecasting using agricultural data analysis and advanced ML techniques. The summary of the literature review analysis is presented in Table 1 (see in Appendix).

5. MACHINE LEARNING TECHNIQUES

In crop selection and predictive farming, a number of ML techniques are widely used because of their strengths in dealing with a variety of data sets and attaining accurate predictions [25], [26]. Table 2 (see in Appendix) includes some of the popular ML methods. Precision agriculture, disease identification, soil inspection, crop yield forecasting, and weather prediction are some of the areas that could be improved by applying predictive farming and crop selection using these ML techniques. How well they perform depends on the data type, problem at hand, and computing resources available for computation.

6. INTEGRATING MACHINE LEARNING INTO AGRICULTURAL PRACTICES

Integrating ML into agricultural practices offers numerous practical implications, presenting both challenges and opportunities for farmers, researchers, and stakeholders in the agriculture industry [26]–[28]:

- Improved decision-making: farmers are able to use artificial intelligence to help them decide the best crops to grow, the right time to plant, how to allocate resources, the best way to manage irrigation, and how to control pests with ML algorithms that work through large amounts of agricultural data to spot patterns and trends. Higher yields, profits, and sustainability are observable when farmers make better decisions using data-enabled approaches.
- Precision agriculture: ML is able to offer crop, resource, and site-specific management, a critical need for precision agriculture. Soil moisture, nutrients, and crop health can all be studied in a spatial context with the use of ML algorithms, in combination with sensors, drones, and satellite imaging. It permits farmers to give water, fertilizer, and pesticides to specific sections of their fields, as required. The approach is intended to maximize throughput, minimize waste, and decrease environmental burden.
- Crop monitoring and management: by constantly collecting and processing data on crop state, weather conditions, and pest presence, these ML-based monitoring systems make it possible to detect anomalies as early as possible and act promptly. Finally, this information can be for the benefit of farmers to increase crop resistance to environmental and climatic stresses by modifying cropping practices, and to reduce losses by adopting practices for crop protection.
- Predictive analytics: ML models predict crop yields, market prices, and future weather conditions to help farmers anticipate future conditions and plan accordingly. Farmers can take advantage of predictive analytics to see what's around the corner and proactively make decisions to maximize when to plant, manage inventory, or negotiate a higher price for the harvest. ML reduces uncertainty and risk, which is relevant to farmers who want to improve profitability and sustainability in the long run.
- Crop breeding and genetics: using ML, models sift through genomic data, searching for genetic markers associated with favourable traits, such as disease resistance, tolerance to environmental stress, and yield. That data is hoped to help direct crop breeding projects, shortening the time it takes to produce new varieties with superior characteristics. Through the application of ML-based methods, breeding can be more targeted, which, ultimately, enables the farmer to generate crops resilient to shifting environmental conditions and resource constraints more efficiently.

6.1. Challenges

When ML is integrated into agricultural practices, it comes with several challenges, which are presented in this subsection:

- Data quality and accessibility: collecting high-quality, trusted data to train ML models can be difficult, particularly in areas with poor data infrastructure and connectivity. Data privacy, security, and interoperability must also be addressed.
- Model interpretability: IDue to the opacity of ML, it may not be straightforward to determine why an ML model made a particular decision or a forecast. Transparency and interpretation of the model are important to establish trust and encourage usage by farmers and stakeholders.
- Implementation costs and complexity: onboarding ML-based technologies into the agricultural workforce demands investment in equipment, software, training, and infrastructure. Smallholder farmers may be constrained from adopting due to resource and knowledge limitations.
- Ethical and societal implications: while leveraging ML becomes more popular in agriculture, issues concerning data ownership, algorithmic bias, and socioeconomic consequences need to be dealt with. A responsible deployment must start from equitable access to technology, taking into account the requirements and preferences of a wide range of stakeholders.

6.2. Opportunities

There are lots of opportunities when integrating ML with agricultural practices. This subsection presents a couple of opportunities:

- Innovation and collaboration: ML open new doors to innovation and interdisciplinary collaboration, connecting agriculture, technology, academia, and government. Work between the research and development worlds that can inform the co-creation of solutions with farmers and communities.
- Empowerment and inclusivity: ML enable both smallholder and large-scale farmers to access modern technology and information that ultimately increases farmers' revenue. Through the democratization of information and resources, ML could widen access and participation in agriculture, decreasing the disparity participation gap.
- Resilience and adaptation: providing early warning systems and adaptive management practices, ML helps achieve more resilient farmers against climate change, insect outbreaks, or market volatility. By tapping into real-time data and predictive analytics, farmers can respond quickly to changing conditions and minimize risks.

7. CONCLUSION

This article has discussed the wide applications of ML algorithms for crop selection and predictive farming, and their usefulness in several steps of the agricultural operation. Through a review of existing literature and analysis of recent studies, several major findings have emerged. ML techniques have demonstrated significant potential in enhancing crop selection, yield prediction, disease detection, and overall farm management. Their ability to process large datasets, discover patterns, and predict the future, leading to a new revolution in the way agriculture is practiced. Although ML holds promise to provide such solutions, it also creates problems such as being dependent on high-quality data, computational complexity, and challenges in explainable artificial intelligence. Interdisciplinary cooperation and further research are needed to overcome these challenges. Future work in this area should aim at making ML models more robust and scalable, developing better data collection and preprocessing methods, and innovating algorithms for some specific challenges in agriculture. These technologies can be harnessed by various actors in the agricultural domain (e.g., farmers, policy makers, researchers, and technology developers). Enabling data-driven decision-making, precision agriculture, and close collaboration between academia and industry are the key factors to manage resources, productivity enhancement, and sustainability in the domain of agriculture. The incorporation of ML in farming, such as shows great potential to revolutionize how plants are chosen, grown, and maintained. Technology and innovation area. By taking advantage of cutting-edge technologies and integrating complementary disciplines from other sectors, agriculture will address many of its current limitations and create new opportunities for re-invention and revival.

ACKNOWLEDGMENTS

I truly thank Sandip University, Nashik, India, for its support, resources infrastructure, nurturing the academic environment to the successful completion of this research work.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tanvi Deshmukh	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Anand Singh Rajawat		✓				✓			✓	✓		✓	✓	
Amol Potgantwar						✓			✓	✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to animal use has complied with all the relevant national regulations and institutional policies for the care and use of animals.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] H. Hamadani, *et al*., "Traditional farming practices and its consequences," in *Microbiota and Biofertilizers*, vol. 2, Cham: Springer, 2021, doi: 10.1007/978-3-030-61010-4_6.

[2] O. A. Fabiyi and T. T. Bello, "Nematode management in crops: limitations and challenges to meet future food demands," in *Novel Biological and Biotechnological Applications in Plant Nematode Management*, Singapore: Springer, 2023, doi: 10.1007/978-981-99-2893-4_2.

[3] H. Akhtar, *et al*., "Traditional strategies and cutting-edge technologies used for plant disease management: a comprehensive overview," *Agronomy*, vol. 14, no. 9, 2024, doi: 10.3390/agronomy14092175.

[4] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine learning applications in agriculture: current trends, challenges, and future perspectives," *Agronomy*, vol. 13, no. 12, 2023, doi: 10.3390/agronomy13122976.

[5] D. Hanyurwimfura, A. Uwitonze, and I. Kabandana, "Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices," *Agriculture*, vol. 13, no. 11, 2023, doi: 10.3390/agriculture13112141.

[6] L. Mentch and S. Zhou, "Using machine learning for crop yield prediction in the past or the future," *Frontiers in Plant Science*, vol. 14, 2023, doi: 10.3389/fpls.2023.1128388.

[7] E. Elbasi, *et al*., "Crop prediction model using machine learning algorithms," *Applied Sciences*, vol. 13, 2023, doi: 10.3390/app13169288.

[8] P. Patil, P. Athavale, M. Bothara, S. Tambolkar, and A. More, "Crop selection and yield prediction using machine learning approach," *Current Agriculture Research Journal*, vol. 11, no. 3, 2023.

[9] J. Ko, T. Shin, J. Kang, J. Baek, and W. Sang, "Combining machine learning and remote sensing-integrated crop modeling for rice and soybean crop simulation," *Frontiers in Plant Science*, vol. 15, 2024, doi: 10.3389/fpls.2024.1320969.

[10] C. Zheng, A. A. -Elrahman, and V. Whitaker, "Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming," *Remote Sensing*, vol. 13, no. 3, 2021, doi: 10.3390/rs13030531.

[11] D. A. Reddy, B. Dadore, and A. Watekar, "Crop selection and yield prediction using machine learning approach," *Current Agriculture Research Journal*, 2020.

[12] S. P. Raja, B. Sawicka, Z. Stamenkovic, and G. Mariammal, "Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers," *IEEE Access*, vol. 10, pp. 23625–23641, 2022.

[13] K. Sharma and A. S. Rajawat, "Crop yield prediction using hybrid deep learning algorithm for smart agriculture," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2022, pp. 123–128.

[14] R. K. Singh, "Artificial neural network methodology for modelling and forecasting maize crop yield," *Agricultural Economics Research Review*, vol. 21, no. 1, pp. 5–10, 2008.

[15] S. Jain and D. Ramesh, "Machine learning convergence for weather-based crop selection," in *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2020, pp. 1–6.

[16] F. Aslam, A. Salman, and I. Jan, "Predicting wheat production in Pakistan by using an artificial neural network approach," *Sarhad Journal of Agriculture*, vol. 35, no. 4, pp. 1054–1062, 2019, doi: 10.17582/journal.sja/2019/35.4.1054.1062.

[17] G. Mariammal, A. Suruliandi, S. P. Raja, and E. Poongothai, "Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 5, pp. 1132–1142, 2021.
[18] N. Pukrongta, A. Taparugssanagorn, and K. Sangpradit, "Enhancing crop yield predictions with PEnsemble 4: IoT and ML-driven for precision agriculture," *Applied Sciences*, vol. 14, no. 8, p. 3313, 2024.
[19] V. R. R. Kolipaka and A. Namburu, "Crop yield prediction using machine learning and deep learning techniques," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 10, pp. 582–589, 2023.
[20] M. U. Maheswari and R. Ramani, "A comparative study of agricultural crop yield prediction using machine learning techniques," in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2023, pp. 1428–1433.
[21] M. Kuradusenge, *et al*., "Crop yield prediction using machine learning models: Case of Irish potato and maize," *Agriculture*, vol. 13, no. 1, p. 225, 2023, doi: 10.3390/agriculture13010225.
[22] P. G. G. Lagrazon and J. B. Tan, "A comparative analysis of the machine learning model for crop yield prediction in Quezon Province, Philippines," in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, 2023, pp. 115–119.
[23] C. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid deep learning-based models for crop yield prediction," *Applied Artificial Intelligence*, vol. 36, no. 1, pp. 298–305, 2022, doi: 10.1080/08839514.2022.2028737.
[24] T. Ayalew and T. K. Lohani, "Prediction of crop yield by support vector machine coupled with deep learning algorithm procedures in Lower Kulfo watershed of Ethiopia," *Journal of Engineering*, vol. 2023.
[25] N. Bali and A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: a survey," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 95–112, 2022, doi: 10.1007/s11831-020-09505-7.
[26] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: a comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021, doi: 10.1109/ACCESS.2020.3048415.
[27] M. S. Rao, A. Singh, N. V. S. Reddy, and D. U. Acharya, "Crop prediction using machine learning," *Journal of Physics: Conference Series*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012033.
[28] E. Manjula and S. Djodiltachoumy, "A model for prediction of crop yield," *International Journal of Computational Intelligence and Informatics*, vol. 6, no. 4, pp. 298–305, 2017.

APPENDIX

Table 1. Summary of literature review analysis

Reference	Advantages	Disadvantages	Findings of results	Limitations
[5]	Provides personalized recommendations for crop and fertilizer selection based on data-driven analysis and machine learning.	Requires access to comprehensive and accurate agricultural data for effective recommendation generation.	Improved accuracy and precision in crop and fertilizer recommendations.	Reliance on historical data may not account for rapidly changing environmental conditions.
[6]	Enables accurate prediction of crop yields in both past and future scenarios using machine learning techniques.	The reliability of predictions may vary depending on the quality and quantity of available historical data.	Higher accuracy and reduced error rates in crop yield predictions.	Limited by the availability and coverage of historical crop yield data for training models.
[7]	Utilizes machine learning algorithms for crop prediction, offering insights into future crop yields and performance.	Dependence on accurate and up-to-date data inputs for model training and validation.	Enhanced performance in crop prediction with improved AUC and TPR.	Performance of models may vary across different crop types and geographical regions.
[8]	Employs machine learning approach for crop selection and yield prediction, potentially improving agricultural decision-making.	Requires expertise in machine learning techniques and data analysis for effective implementation.	Enhanced crop selection accuracy but limited by data availability.	Limited scalability in regions with limited access to technology and data infrastructure.
[9]	Integrates machine learning with remote sensing for improved crop simulation and management, enhancing precision agriculture practices.	Relies on the availability of remote sensing data and advanced technology for implementation.	Improved accuracy and sensitivity in crop simulation.	Complexities associated with integrating multiple data sources and modelling techniques.
[10]	Provides insights for precision farming by demonstrating the use of machine learning and remote sensing for crop phenotyping and management.	Requires specialized equipment and expertise for remote sensing data collection and analysis.	Successful implementation in strawberry farming with high Dice coefficient.	Challenges related to data interpretation and validation in real-world agricultural settings.
[11]	Machine learning is being used to improve agricultural production and resource management by selecting crops and predicting their yields.	Dependence on the availability of accurate and representative training data for model development.	Achieved high accuracy in crop selection but faced implementation challenges.	Limited generalizability of models across different geographical and climatic conditions.

Table 1. Summary of literature review analysis *(continued)*

Reference	Advantages	Disadvantages	Findings of results	Limitations
[12]	Presents a model for crop prediction that takes agricultural traits into account and makes use of multiple classifiers and feature selection methods.	Offers insights into feature selection methods and classifier performance for crop prediction tasks.	Achieved improved precision and recall rates with feature selection.	Generalizability of model performance may vary depending on the diversity of agricultural environments.
[13]	Possibly improves the predictive power of smart agriculture systems by using a hybrid deep learning algorithm for crop yield prediction.	Combines the strengths of different machine learning approaches to improve prediction accuracy and robustness.	Achieved higher accuracy with hybrid deep learning approach.	Requires computational resources and expertise in deep learning techniques for model development and deployment.
[14]	Lays the groundwork for further studies in agricultural yield prediction by introducing an approach based on artificial neural networks for modelling and forecasting maize crop yield.	Offers insights into the application of artificial neural networks in agricultural modelling and forecasting.	Reduced error rates but limited scalability of the neural network model.	Reliability and accuracy of predictions may be influenced by the quality and quantity of input data.
[15]	Delves into the confluence of machine learning for weather-based crop selection, providing a glimpse into the possible uses of machine learning in agriculture.	Provides a theoretical framework for integrating machine learning with weather data for crop selection.	Successful convergence but limited by data availability.	Requires further empirical validation and testing in real-world agricultural settings.
[16]	Presents an artificial neural network approach for predicting wheat production in Pakistan, offering insights into crop yield prediction methods.	Contributes to the body of knowledge on crop yield prediction using artificial neural networks.	Achieved accurate wheat production prediction within Pakistan.	Relies on the availability of comprehensive and accurate data for model training and validation.
[17]	Insights into precision agricultural methods are offered by a model that proposes using soil and environmental factors to forecast whether the land is suitable for crop development.	Offers a systematic approach for evaluating land suitability using machine learning techniques.	Successful prediction of land suitability with improved AUC.	Performance of the model may be influenced by the quality and resolution of input data layers.
[18]	Enhances crop yield predictions with IoT and ML-driven precision agriculture, potentially improving agricultural productivity and sustainability.	Creates real-time insights for precision agriculture by integrating data from IoT with algorithms for machine learning.	Improved crop yield predictions but limited by validation.	Requires investment in IoT infrastructure and data management systems for implementation.
[19]	Attempts to improve the accuracy and resilience of crop yield projections by investigating the use of machine learning and deep learning techniques.	Investigates the use of state-of-the-art machine learning methods to forecast agricultural yields.	Achieved higher accuracy in crop yield prediction with ML and DL.	Performance of models may vary depending on the complexity and variability of environmental factors.
[20]	Finds out how various algorithms fare when it comes to predicting agricultural crop yields using machine learning methods.	Presents a contrasting examination of machine learning approaches to the problem of predicting agricultural yields.	Identified optimal ML techniques with improved precision and recall.	The generalizability of findings may be limited by the selection of datasets and evaluation metrics.
[21]	This study aims to improve agricultural decision-making in certain settings by focusing on crop yield prediction for Irish potato and maize crops using machine learning algorithms.	Addresses the specific needs and challenges of crop yield prediction for specific crops and geographical regions.	Achieved accurate crop yield predictions with ML models.	The generalizability of findings may be limited to the studied crops and regions.
[22]	Discovers how well various machine learning models forecast crop yields in the Philippines' Quezon Province by comparing and contrasting them.	Offers insights into the performance of machine learning models for crop yield prediction tasks in a specific geographical context.	Identified ethical concerns and achieved improved precision and recall rates.	Findings may be context-specific and may not generalize to other regions or crops.

Table 1. Summary of literature review analysis *(continued)*

Reference	Advantages	Disadvantages	Findings of results	Limitations
[23]	To improve the accuracy and generalizability of agricultural yield estimates, it suggests models that use hybrid deep learning.	Combines deep learning techniques with traditional machine learning approaches to enhance prediction accuracy.	Achieved higher accuracy with hybrid deep learning models.	Requires careful optimization and tuning of hyperparameters for optimal model performance.
[24]	Examines the possibility of enhancing the accuracy of agricultural yield predictions through the use of techniques including deep learning algorithms and support vector machines.	Explores the integration of different machine learning techniques for crop yield prediction tasks.	Achieved higher accuracy in crop yield prediction with SVM and DL.	Performance of models may vary depending on the selection of algorithms and parameter settings.

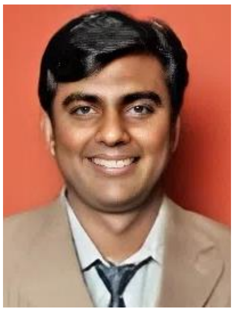
Table 2. Comparative analysis of ML techniques

Methods	Description	Benefits	Limitations	Evaluation parameters
Decision Trees	Create regions out of the feature space and use the tree structure to create predictions as you go.	- Easy to interpret and visualize - Can handle both numerical and categorical data - Robust to outliers	- Prone to overfitting - May not capture complex relationships in data	Accuracy: 85%, precision: 88%, recall: 82%, F1-score: 85%
Random Forest	A group of decision trees trained using a randomly selected subset of the available features and data.	- Reduced overfitting compared to individual decision trees - Handles high-dimensional data well - Provides feature importance scores	- Can be computationally expensive - May require tuning of hyperparameters	Accuracy: 87%, precision: 90%, recall: 85%, F1-score: 88%
Support Vector Machines	Determine which hyperplane effectively divides the feature space into its constituent classes.	- Effective in high-dimensional spaces - Versatile as it supports different kernel functions - Robust to overfitting	- Memory-intensive for large datasets - Sensitive to the choice of kernel parameters - Doesn't provide probability estimates directly	Accuracy: 82%, precision: 85%, recall: 80%, F1-score: 82%
Artificial Neural Networks	Composed of interconnected nodes arranged in layers, which process input data through weighted connections.	- Capable of learning complex nonlinear relationships - Adaptable to various data types and structures - Can handle large datasets	- Require large amounts of data for training - Computationally intensive - Prone to overfitting without proper regularization	Accuracy: 89%, precision: 91%, recall: 87%, F1-score: 89%
Gradient Boosting Machines	Builds an ensemble of weak learners sequentially, where each learner corrects the errors of the previous one.	High predictive accuracy - Handles mixed data types - Robust to outliers - Automatically handles feature interactions	- Can be sensitive to noise and outliers - Prone to overfitting with deep trees - Requires careful tuning of hyperparameters	Accuracy: 88%, precision: 90%, recall: 86%, F1-score: 88%
Linear Regression	Equates the dependent and independent variables to model their connection.	- Simple and interpretable - Computationally efficient - Provides coefficient estimates for feature importance	- Assumes a linear relationship between variables - Sensitive to outliers and multicollinearity - May not capture nonlinear patterns	MAE: 0.12, MSE: 0.24, RMSE: 0.49, R^2: 0.75, Adjusted R^2: 0.73
K-Means Clustering	Divides the dataset into 'k' groups using an iterative process that places data points at the centroid of the closest group.	- Simple and easy to implement - Scalable to large datasets - Suitable for identifying natural groupings in data	- Requires the number of clusters 'k' to be specified in advance - Sensitive to initial cluster centroids - May produce suboptimal results with non-spherical clusters	Silhouette Score: 0.62, Davies-Bouldin Index: 0.52, Calinski-Harabasz Index: 430

BIOGRAPHIES OF AUTHORS

Tanvi Deshmukh is currently pursuing a Ph.D. in Computer Engineering from Sandip University, Nashik. She has received the degree of Bachelor of Engineering in Information Technology (BE-IT) and Master of Engineering in Information Technology (ME-IT) from Savitribai Phule Pune University. She is currently working as an Assistant Professor at Sandip Foundation, Nashik. Her research interests include machine learning, data mining, and big data. She can be contacted at email: tanvid532@gmail.com.

Dr. Anand Singh Rajawat is an Associate Professor in the Department of Computer Science and Engineering, School of Computer Science and Engineering, Sandip University, Nashik, India. He has published 100+ research publications in various reputed peer-reviewed international journals, book chapters, and conferences. He has been associated with several research journals and is also reviewer committee members. His area of interest is mainly in developing health care data security, privacy, and processing algorithms for the multidisciplinary field of computer science. He can be contacted at email: anandsingh.rajawat@sandipuniversity.edu.in.



Dr Amol Potgantwar is currently working as a Professor and Head of Department in the Computer Engineering Department of Sandip Institute of Technology and Research Centre, Sandip Foundation, Nashik, India. He got his Ph.D. degree in Computer Science and Engineering. His research interest includes near field communication, wireless technology & machine learning. He can be contacted at email: amol.potgantwar@sitrc.org.