JOURNAL OF MULTIDIMENSIONAL RESEARCH & REVIEW

http://www.jmrr.org Volume: 6, Issue: 1, April 2025 | ISSN: 2708-9452

Agriculture Crop Yield Prediction Using Machine Learning

Santha M

Assistant Professor,PG & Research Department of Computer Science and Applications
Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode,
Namakkal, Tamilnadu, India.

Indhumathi P

MCA Student, PG& Research Department of Computer Science and Applications
Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode
Namakkal, Tamilnadu, India.

Abstract

Agriculture, the principal sector of global food production, must feed an increasing population while also adapting to changing environmental conditions. In an increasingly uncertain world, this technological marriage of data science and agriculture provides a feasible solution to ensuring a sustainable food supply and improving farming practices. We proposed employing Support Vector Machines (SVM) as a crucial tool for agricultural yield prediction. SVM is a powerful machine learning algorithm that excels at detecting complex patterns in datasets, particularly those related to agricultural yields and environmental factors. Farmers who use SVM may gain valuable knowledge about agricultural productivity, allowing them to make informed decisions in the face of climate change uncertainty. Crop yield prediction models can help farmers make better judgments regarding their crops, highlighting SVM's potential as a uniquely successful tool.

Keywords: Crop Yield Prediction, Decision Support System, Precision Agriculture

---

1 Introduction

Machine learning (ML) techniques are used in a wide range of industries, including grocery stores, to forecast user phone usage and assess consumer behavior. Agriculture has been employing machine learning for some years. One of the most difficult aspects of precision agriculture is forecasting crop yield, and numerous models have been proposed and proven useful thus far. To answer this problem, many datasets must be employed because crop yield is impacted by a variety of factors such as soil, weather, fertilizer use, and seed type. This shows that estimating agricultural production is a complex process.

1.1 Crop Yield Prediction

Machine learning (ML) techniques are used in a wide range of sectors, including grocery shops, to estimate user phone usage and analyze consumer behavior. Agriculture has been employing machine learning for some years. One of the most difficult aspects of precision agriculture is forecasting crop yield, and numerous models have been proposed and proven useful thus far. To answer this problem, many datasets must be employed because crop yield is impacted by a variety of factors such as soil, weather, fertilizer use, and seed type. This shows that estimating agricultural production is a complex process. Even though agricultural yield prediction algorithms can now reliably estimate actual yield, greater yield prediction performance remains preferable.

1.2 Decision Support System

A decision support system (DSS) is a type of computer software that helps with choices, evaluations, and actions inside a business or organization. A DSS sorts and analyzes massive data sets before compiling specific information for use in decision-making and issue solving. A DSS frequently employs objective or expected revenue, historical sales data from multiple time periods, and data on inventory or activities. A decision support system collects, analyzes, and synthesizes data to provide comprehensive information reports. This is how a DSS differs from a standard operations application, which just collects data. The DSS can be completely automated or conducted solely by humans. It might occasionally combine the two. The finest systems analyze data and act on behalf of the user. They allow human users to make decisions faster and with better knowledge, at the very least.

1.3 Precision Agriculture

Precision agriculture is a revolutionary agricultural technique that employs cutting-edge technology to optimize resource allocation, increase efficiency, and maximize yields while minimizing environmental impact. Precision agriculture, at its core, makes use of a wealth of data from satellites, drones, and sensors to precisely tailor agricultural operations to individual plants or small areas of fields. Farmers may utilize this information to make informed decisions regarding irrigation, fertilization, pest control, and crop selection, which will result in cheaper input costs, improved crop quality, and more sustainable farming practices.

2 Literature Review

In this research, Alexandros proposed an application of deep learning to the problem of agricultural production forecast. However, there has been a lack of systematic analysis in previous studies. Hence, the goal of this study is to offer an overview of the state-of-the-art application of deep learning in agricultural production prediction. To accomplish this, a systematic literature review (SLR) was conducted to identify and evaluate the most pertinent research. Out of the 456 relevant publications identified, 44 primary studies were chosen for further analysis using selection and quality rating criteria. The primary studies were thoroughly analyzed and synthesized in terms of key motivations, target crops, algorithms, features, and data sources. The data showed that the Convolutional Neural Network (CNN) algorithm is the most commonly used and performs best in terms of Root Mean Square Error (RMSE). One of the key issues found is a lack of big training datasets, which raises the danger of overfitting and, as a result, reduced model performance in practical applications.

Before developing their own models [**2**], researchers must first understand the challenges in this field. Meanwhile, practitioners face a variety of challenges in developing new crop production prediction models, as described in this SLR research. These issues include the selection of model parameters and methodology, which necessitates a thorough review of existing research. To conduct a thorough evaluation, we reviewed 456 relevant studies. Notably, no systematic literature evaluation has been conducted on the use of deep learning in agricultural production prediction thus far. Oilseed crops are commonly found in fragile agricultural systems that rely on climatic conditions, notably in semi-arid areas. The groundnut (peanut) crop, unlike many other oilseed crops, is extremely vulnerable to several insect and disease attacks. Australia, India, and the United States have all had significant crop losses as a result of these infections. Thrips species, among pests, provide a complex threat to the groundnut crop beginning with the vegetative stage and continuing until harvest. Ramesh Babu Palepu [**3**]et al. presented a system in which agriculture plays the primary role in meeting the world's food demands. Agriculture is recognized as the economic backbone of many developing nations, including India. The application of data mining tools in agriculture, particularly soil analysis, has the potential to improve decision-making and increase agricultural yields. Soil analysis is critical for addressing a wide range of agricultural issues. This research looks at the use of data mining in soil analysis in agriculture, as well as the contributions of other authors on the issue. Data mining techniques are extremely beneficial in the field of soil analysis. Although data mining is widely used in a variety of sectors, and there are several off-the-shelf tools and approaches available, the use of data mining in agricultural soil databases is still a relatively new research topic. Nowadays, data mining concepts and techniques are used to solve agricultural problems. This study investigates the use of data mining techniques in agriculture.

The technique proposed by V. Rajeswari et al. emphasizes the importance of soil in agriculture and attempts to predict soil type using data mining classification techniques. The study uses JRip, J48, and Naive Bayes algorithms to extract information from soil data, focusing on red and black soil types. The statistics suggest that the JRip model produces more consistent results and increases soil type prediction accuracy. The report also emphasizes the use of data mining in agriculture for soil classification, wasteland management, and crop and pest management. Other research has focused on data mining tools for anticipating soil fertility rates, finding agricultural information, and calculating agricultural production analysis. Efficient data mining algorithms may be developed to increase the accuracy of categorization in large soil data sets while simultaneously addressing Big Data problems.

Nabila Chergui [**5**] et al. contend in their system that recent improvements in information and communication technology have had a significant impact on several sectors of the global economy. The widespread availability of digital devices, as well as advancements in artificial intelligence and data analytics, have all led to the growth of digital agriculture. This new approach to agriculture has resulted in revolutionary strategies that boost productivity and efficiency while emphasizing environmental sustainability. The use of powerful digital equipment and data science has enabled the collection and analysis of large agricultural datasets, giving farmers, agronomists, and professionals a better understanding of farming operations and the ability to make more informed decisions. This paper conducts a detailed assessment of data mining technologies, particularly in the context of digital agriculture. The focus of this study is on agricultural yield management and monitoring, as we examine the many data mining methodologies employed in this industry. Furthermore, we examine a number of existing studies that apply data analytics in agriculture.

3 Existing System

Crop production prediction using crop, soil, water, and environmental factors is one example of such research. Deep learning models are frequently used to extract and forecast crucial agricultural properties. Despite the fact that these techniques may handle the yield prediction problem, they have the following limitations: unable to give a direct linear or nonlinear mapping between crop production values and raw data; moreover, the efficacy of those models is heavily reliant on the quality of the features obtained. Despite the aforementioned shortcomings, deep reinforcement learning provides advice and motivation. Deep reinforcement learning combines the intelligence of deep learning with reinforcement learning to develop a complete framework for forecasting crop output, allowing raw data to be translated into crop prediction values. To forecast agricultural yield, the current study constructs a Deep Recurrent Q-Network model, which combines a recurrent neural network deep learning algorithm with the Q-learning reinforcement learning method. The data parameters drive the recurrent neural network's increasingly stacked layers. Using the input parameters, the Q-learning network creates a crop production prediction environment. A linear layer maps the recurrent neural network's output values to Q-values.

3.1 Existing Algorithm Description

The algorithm described in the abstract combines deep learning and reinforcement learning to predict crop yield more effectively. It uses a Deep Recurrent Q-Network (DRQN), which integrates a Recurrent Neural Network (RNN) with the Q-Learning reinforcement learning algorithm. The RNN processes sequential data by stacking multiple layers, allowing the model to capture temporal dependencies from the input parameters such as environmental, soil, water, and crop-related factors. The output of the RNN is then mapped to Q-values through a linear layer. These Q-values represent the expected rewards for different actions in the prediction environment. The reinforcement learning agent uses a combination of parametric features and threshold-based decision making to predict crop yield. Through iterative training, the agent learns to minimize prediction errors and improve forecasting accuracy. This approach allows the model to directly map raw input data to yield predictions, overcoming the limitations of traditional deep learning models. The integration of reinforcement learning also helps the model adapt to dynamic agricultural conditions.

3.2 Deep Recurrent Q-Network (DRQN)

- It combines Recurrent Neural Networks (RNNs) and Q-Learning.
- RNN is a type of deep learning model good at handling sequential data (like time series very important for crops that grow over time).
- Q-Learning is a reinforcement learning technique where an agent learns to make decisions by maximizing a reward (here, accurate yield prediction)

4 Proposed System

To enhance crop production forecasts in agriculture, the proposed strategy includes cutting-edge data science approaches, including support vector machines (SVM). Starting with the load data module, the system collects detailed datasets on climate, soil parameters, and previous crop yields. The data preparation module next cleans up the data by filling in missing values and standardizing the characteristics, preparing it for analysis. The feature selection module optimizes the model's performance by identifying key variables. The "training and testing" module teaches and assesses the SVM algorithm to ensure that it can adapt to a wide range of agricultural settings. Finally, the crop yield prediction using the SVM module uses the trained model to generate precise and trustworthy forecasts, providing farmers with vital information for making educated crop management decisions.

4.1 Load Data

The system imports and receives the necessary agricultural datasets in the Load Data module. This requires gathering information on climatic variables, soil properties, previous crop yields, and other relevant environmental elements. The module ensures that the supplied data is accurate, full, and indicative of the agricultural scenarios under consideration. The crop yield prediction system's subsequent modules are based on the foundation formed by the successful loading of data.

4.2 Data Preprocessing

The module in charge of data preprocessing is responsible for honing and loading the data for analysis. This includes removing outliers, filling in missing data, and scaling or normalizing features to ensure consistency and compatibility with the Support Vector Machines (SVM) approach. Data preprocessing is critical for enhancing the accuracy and reliability of predictions because it makes it feasible for the SVM model to efficiently discover patterns in the input data and learn from them.

4.3 Feature Selection

During the feature selection stage, relevant variables from the preprocessed data are chosen to improve the performance of the SVM model. The purpose of this step is to identify the important factors that have the greatest influence on crop output forecasts. By selecting a subset of key characteristics, the model's computational efficiency improves while superfluous or unnecessary data is removed, resulting in a more efficient and successful prediction process.

4.4 Training and Testing

In the training and testing module, the SVM algorithm is put to use. A subset of the pre-processed data is used to train the model, which learns how input characteristics and crop yields are related. To ensure that the model is generalizable, its performance is next evaluated using a testing set of data, which is a separate collection of data from the training set. This module modifies the SVM parameters and assesses the model's dependability and accuracy in predicting crop yields under various circumstances.

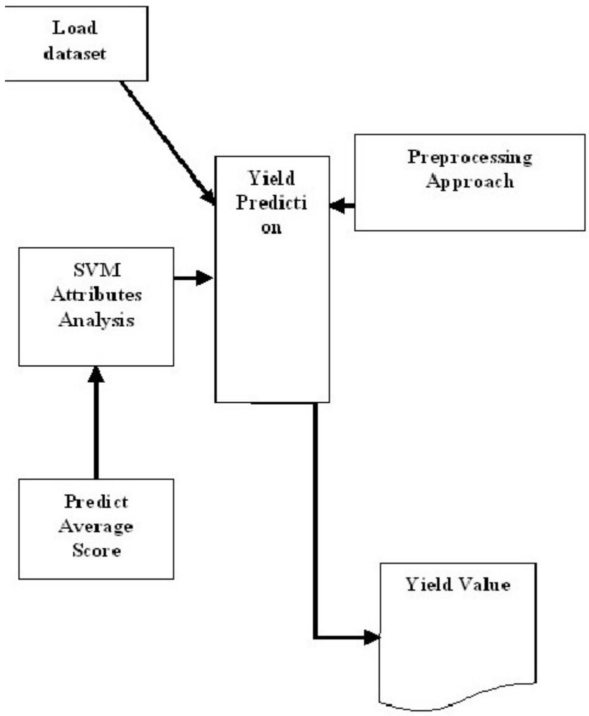

Figure 1: Block Diagram

The image depicts a flowchart illustrating a process for predicting yield using a machine learning approach. It begins with loading a dataset, which is then input into the Yield Prediction module. Before prediction, the data undergoes a Preprocessing Approach to prepare it for analysis, ensuring the data quality and relevance. The preprocessed data feeds into the yield prediction model. Simultaneously, an SVM (Support Vector Machine) Attributes Analysis is conducted to select and analyse important features from the dataset, which further informs and enhances the yield prediction. After prediction, the process continues to predict an average score, likely assessing model performance or confidence, which loops back to refine the SVM attribute analysis. Finally, the yield prediction outputs a Yield Value, which is the final result of the process. Overall, the flowchart outlines a systematic approach combining preprocessing, feature analysis, prediction, and evaluation to accurately estimate yield outcomes.

4.5 Crop Yield Prediction Using SVM

Crop Yield Prediction The trained model is applied to fresh, previously unseen data using the SVM module to create predictions. Farmers can obtain crop yield projections by entering pertinent environmental data. The SVM algorithm provides reliable and accurate forecasts by recognizing patterns, providing farmers with essential information to assist them in making crop management decisions. This seminar promotes proactive and educated agricultural practices by showcasing the real-world use of an SVM-based crop production forecast system.

5 Algorithm Details

Support Vector Machine (SVM) is a supervised machine learning technique that is commonly used in classification and regression issues. We represent each data item as a point in n-dimensional space, with the value of each feature representing the value of a specific coordinate.
Step 1: The total training set is again divided into two different sets. (train and holdout)
Step 2: Train these elected base models with the first part (train).
Step 3: Test them with these cond parts. (holdout)
Step 4: Now, the prediction test results are used to measure the accuracy.
There are several alternative hyperplanes for separating the two types of data points. Our goal is to determine the plane with the greatest margin, which is the greatest distance between data points from both classes. Maximizing the margin distance gives some reinforcement, allowing subsequent data points to be categorized with more certainty. Hyperplanes are decision boundaries used to categorize data points. Data points on each side of the hyperplane can be assigned to various classifications. In addition, the number of features determines the hyperplane's dimension. If the number of input features is two, the hyperplane is simply a line. If there are three input characteristics, the hyperplane becomes two-dimensional. It's impossible to conceive when the number of characteristics reaches three. Support vectors are data points near to the hyperplane that impact its location and direction. We use these support vectors to maximize the classifier's margin.

6 Result Analysis

The performance of the current agricultural crop recommendation algorithm, which achieves 75% accuracy, is compared to the proposed algorithm, which improves significantly to 81% accuracy. It is apparent that the newly introduced technology has better predictive capabilities. The suggested system, which may use SVM and a large dataset including seasonal elements, outperforms the current approach by providing more exact crop predictions. This improvement in accuracy shows that the suggested method has the potential to significantly improve decision-making for farmers and agricultural specialists, resulting in optimal crop selection and increased overall agricultural output.
The existing algorithm achieves an accuracy of 75%, while the proposed algorithm shows

a higher accuracy of 81%. This indicates that the proposed algorithm outperforms the existing one by 6 percentage points, suggesting it may be more effective or efficient in the given task. The improvement in accuracy highlights the potential benefits of implementing the proposed algorithm over the existing one.

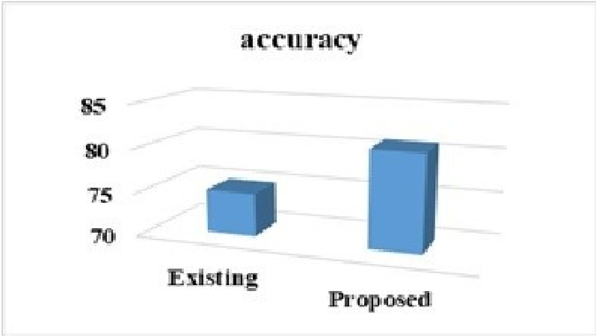

Figure 2: Comparison Graph

The image shows a 3D bar graph comparing the accuracy between an existing system and a proposed system. The vertical axis represents accuracy values, ranging from 70 to 85. The bar corresponding to the existing system reaches slightly above 75, while the bar for the proposed system rises significantly higher, close to 82 or 83. This visual representation clearly indicates that the proposed system achieves a noticeably higher accuracy than the existing system, suggesting that improvements or new methodologies introduced in the proposed approach have enhanced overall performance.

7 Output

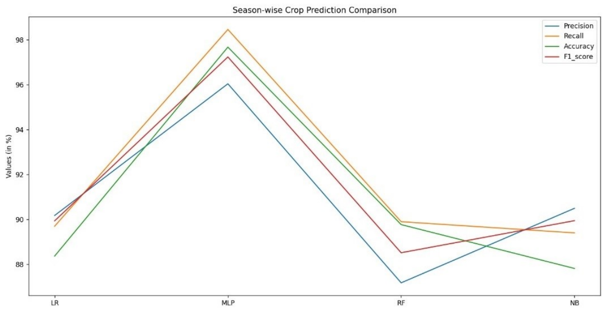

Figure 3: Output

The image shows a line graph comparing the performance of different machine learning models—Logistic Regression (LR), Multi-Layer Perceptron (MLP), Random Forest (RF), and Naive Bayes (NB)—for season-wise crop prediction. The performance metrics considered are Precision, Recall, Accuracy, and F1-score, all measured in percentages. The graph indicates that MLP achieves the highest values across all metrics, peaking around 98%, suggesting superior prediction capability compared to the other models. Logistic Regression shows moderate performance, with values around 90–92%. Random Forest sees a noticeable drop across all metrics, falling below 90%. Naive Bayes performs slightly better than Random Forest in some metrics but generally stays lower compared to MLP.

Overall, the graph highlights that MLP is the most effective model for crop prediction in this comparison, outperforming the others significantly.

8 Conclusion

Finally, the use of Support Vector Machines (SVM) in crop production prediction is a practical and successful technique for assisting farmers in dealing with changing agricultural challenges. By using the modules for data loading, preprocessing, feature selection, and training/testing, the proposed system successfully leverages SVM's strengths to forecast agricultural yields. The service supports farmers in optimizing crop management by providing practical insights based on substantial data. Beyond its immediate benefits, this technology fusion has the potential to promote sustainable agriculture practices in an increasingly unpredictable environment. As a result, the proposed technique represents a significant development in precision farming that will help fight for global food security while also tackling the dual concerns of feeding a rising population and reducing the consequences of climate change. Subsequent study in this area will include developing and strengthening the proposed crop yield prediction model. Subsequent investigations may look into the use of sophisticated machine learning algorithms or ensemble techniques to improve the model's forecast precision and resilience. Furthermore, including data streams and cutting-edge sensors can increase the system's dynamic flexibility and make agricultural decision-making more responsive.

References

[1] S. Li, S. Peng, W. Chen, and X. Lu, "INCOME: Practical land monitoring in precision agriculture with sensor networks," Comput. Commun., vol. 36, no. 4, pp. 459–467, Feb. 2013.

[2] A. D. Jones, F. M. Ngure, G. Pelto, and S. L. Young, "What are we assessing when we measure food security? A compendium and review of current metrics," Adv. Nutr., vol. 4, no. 5, pp. 481–505, 2013.

[3] G. E. O. Ogutu, W. H. P. Franssen, I. Spit, P. Omondi, and R. W. Hutjes, "Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts," Agric. For. Meteorol., vols. 250–251, pp. 243–261, Mar. 2018.

[4] M. E. Holzman, F. Carmona, R. Rivas, and R. Niclòs, "Early assessment of crop yield from remotely sensed water stress and solar radiation data," ISPRS J. Photogramm. Remote Sens., vol. 145, pp. 297–308, Nov. 2018.

[5] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," Trends Plant Sci., vol. 21, no. 2, pp. 110–124, 2016.

[6] R. Whetton, Y. Zhao, S. Shaddad, and A. M. Mouazen, "Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI," Comput. Electron. Agric., vol. 138, pp. 127–136, Jun. 2017.

[7] Y. Dash, S. K. Mishra, and B. K. Panigrahi, "Rainfall prediction for the Kerala state of India using artificial intelligence approaches," Elect. Eng., vol. 70, pp. 66–73, Aug. 2018.

[8] W. Wieder, S. Shoop, L. Barna, T. Franz, and C. Finkenbiner, "Comparison of soil strength measurements of agricultural soils in Nebraska," J. Terramech., vol. 77, pp. 31–48, Jun. 2018.

[9] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, and Z. Li, "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," Remote Sens. Environ., vol. 210, pp. 35–47, Jun. 2018.

[10] X. E. Pantazi, D. Mosholu, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," Comput. Electron. Agric., vol. 121, pp. 57–65, Feb. 2016.

Cite this article:

Santha M & Indhumathi P, “Agriculture Crop Yield Prediction Using Machine Learning”, Journal of Multidimensional Research and Review (JMRR), Vol.6, Iss.2, pp.69-78, 2025