

A Novel Hybrid Method for Predicting Specific Crop Yield

B.Rebecca^{1*}, P.Siva Padmini², V.Dhanakodi³, M.Gayathri⁴

¹Department of Computer Science and Engineering, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad, Telangana, India

²Department of Computer Science and Engineering-Data Science, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad, Telangana, India

^{3,4}Department of Computer Science and Engineering, Mahendra College of Engineering, Minnampalli, Tamil Nadu, India

Received: 02 June 2024 • Accepted: 03 June 2024 • Published Online: 04 June 2024

Abstract: Predicting crop yields is crucial for economic planning, food security, and agricultural management. Conventional approaches frequently need accurate results regarding the complex interaction between environmental conditions and crop-specific traits. By combining the best features of machine learning algorithms with expert-level domain knowledge, we provide a new hybrid approach to predicting crop yields that is both state- and crop-specific. Our approach integrates historical yield data, satellite imagery, meteorological data, soil properties, and crop-specific information to build robust predictive models. Specifically, we employ ensemble learning techniques, including random forest and gradient boosting, to capture complex nonlinear relationships and improve prediction accuracy. Additionally, we incorporate domain knowledge through feature engineering and selection to enhance model interpretability and generalization. We validate our method using real-world datasets from diverse geographical regions and crops, demonstrating its superior performance compared to traditional approaches. Our proposed hybrid method offers a promising solution for accurate and reliable crop yield prediction, facilitating informed agricultural decision-making.

Key words: Machine learning, Hybrid method, Ensemble learning, Satellite imagery, Agricultural management, Crop yield prediction.

*Correspondence: Associate Professor, Department of Computer Science and Engineering, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad, Telangana, India. Email: baderebecca@yahoo.co.in <https://doi.org/10.58599/IJSMCSE.2024.1109>
Vol. 1, No. 1, June 2024, pp:32-38

1. Introduction

Agriculture provides sustenance, jobs, and economic support to many people worldwide. Providing enough food has become urgent because the global population is predicted to exceed 9 billion by the year 2050. Because it aids farmers, lawmakers, and other stakeholders in making decisions on crop management, distribution of resources, and market dynamics, accurate crop yield prediction is critical to this endeavor. Statistical approaches, expert knowledge, and empirical models were the mainstays for agricultural yield prediction for a long time. However, these methods often provide inaccurate and unreliable results because they must account for the complex relationship between climate variables, crop-specific traits, and management techniques. Recent advancements in technology, including remote sensing, machine learning, and big data analytics, have the potential to transform agricultural forecasting. These tools offer unprecedented opportunities to enhance the precision and scalability of crop yield prediction, harnessing vast amounts of data from diverse sources. For example, satellite imagery provides valuable insights into crop health, growth patterns, and environmental conditions at various spatial and temporal scales. Meteorological data offers critical information on temperature, precipitation, and humidity, influencing crop development and yield potential. By integrating these datasets with machine learning algorithms, we can develop sophisticated predictive models capable of capturing nonlinear relationships and adapting to dynamic environmental conditions.

While machine learning holds promise for crop yield prediction, its effective application in real-world agricultural systems is not without challenges. The heterogeneity of agricultural systems, characterized by diverse crop types, climatic zones, soil types, and management practices, poses a significant obstacle. Developing generic models that perform well across these diverse contexts is inherently difficult due to the inherent variability and complexity of agricultural systems. This underscores the need for state- and crop-specific approaches that tailor predictive models to local conditions and crop characteristics. By incorporating domain knowledge, historical data, and advanced analytics techniques, such approaches aim to improve prediction accuracy, robustness, and interpretability. In this study, we address the need for state- and crop-specific crop yield prediction by proposing a novel hybrid method that combines machine learning algorithms with domain-specific knowledge. Our approach leverages historical yield data, satellite imagery, meteorological data, soil properties, and crop-specific information to build accurate and reliable predictive models. We aim to comprehend the complex interplay between climatic factors and agricultural yields by employing ensemble learning techniques such as random forest and gradient boosting. Our models are made more generalizable and more accessible to grasp across different states and types of crops by using selection and feature engineering strategies. Through rigorous validation using real-world datasets, we demonstrate that our proposed approach enhances the accuracy of crop output predictions and enables farmers to make informed decisions.

2. Related Works

This research [1] presents a novel approach to agricultural yield prediction, utilizing data from public sources. By employing machine learning algorithms such as random forest (RF) and support vector machine (SVM), we aim to identify the optimal fertilizer for each crop type. Our primary objective is to develop a robust model that can accurately forecast agricultural yields in the future. In conclusion, our findings demonstrate the potential of machine learning in predicting agricultural yields with a reasonable degree of accuracy. To improve the precision of their yield predictions came up with a method that predicts a suitable yield while also considering the local climate and soil [2]. The goal of this study remains to develop and release a Python-based organization that uses calculated thinking to predict, given a set of inputs, the harvest that will maximize profit while reducing input costs. The long short-term memory (LSTM) and recurrent neural network (RNN) algorithms handle deep learning in this study. In contrast, the support vector machine (SVM) algorithm handles machine learning and integrates state-of-the-art regression techniques into its research, including stacking regression, Kernel Ridge, and Lasso algorithms. The estimate they get from layered regression is substantially less accurate than the one they get from applying the models individually. Now that the results are accessible through a web app, further advances might see the system translated into the farmers' native language and made mobile devices compatible so that they can use it. In this study [3], we investigated the accuracy of various machine learning methods for harvest estimation. As part of the big data computing paradigm, we explored the use of machine learning to predict farmers' output. One of the key measures we considered was the root-mean-square error. While our focus was on predictive machine learning algorithms, we also examined how big data strategies could enhance their prediction capabilities. To address this, we proposed a theoretical framework, which was implemented in a similar fashion.

A data mining analytical approach was developed in [4] to aid farmers in determining soil conditions. The approach gives soil quality assessments a lot of weight since they help increase crop yield with fertilizer recommendations and predict cultivatable crops based on soil type. The system uses managed and unsupervised machine learning (ML) techniques to produce the most accurate result. We will compare the two algorithms' outputs and use the one with the higher confidence level in its results to determine which one is more trustworthy. A unique approach to forecasting greenhouse crop yields using RNNs and temporal convolutional networks (TCNs) [5, 6]. This technique used these two neural networks. The proposed method was evaluated thoroughly using multiple datasets derived from a different real-world greenhouse environment for tomato cultivation. When predicting crop yields, we find that the proposed methodology outperforms both traditional deep neural networks and machine learning methods [7]. The methodology for crop yield calculation is used to compare the actual yields with the projected yields, demonstrating this. Also, regardless

of the desired level of accuracy, the experimental analysis shows that historical yield data is the most significant part of crop production predictions [8]. A model that is enhanced by applying deep learning techniques [9]. The tool predicts crop yields and gives exact details on how much and what kinds of soil elements are needed [10]. Compared to the one that is currently in use, its precision is far better. Analyzing the data provides farmers with a forecast of their harvest, which leads to a rise in their income.

3. Methodology

To progress the accuracy of crop calculation systems, this work uses a statistical method that combines the SVM, RNN, and LSTM algorithms. The forecast as a whole obtains a unique quality from each of the techniques that are used. Separate analyses determine which method produces the most favorable outcomes [10, 11], creating a combined algorithm called proposed hybrid method that comprises the final phase. Some of the fruits are made by various states in India [12, 13]. You will be able to find their support price and ratio of nitrogen, phosphorus pentoxide, and potassium oxide, which are referred to as n-soil, K-soil, and P-soil according to their respective designations. The crop's details and eco-friendly elements like temperature, pH, and humidity are all crucial considerations [14].

Support vector machines (SVMs) have emerged as a powerful tool in regression and classification, a technology that was initially proposed by the researchers. These machines, learning from examples of buildings with low danger, are designed to achieve a non-linear quality through the use of kernel functions. The support vector regression (SVM) method, often employing radial basis functions, can even predict harvest yields. Support vector machines (SVMs) can execute various operations, such as classification and regression, in environments with high or infinite dimensions because of the hyperplanes they generate. Regression and classification are two examples of such problems. Since broader edges are frequently associated with fewer classifier speculative mistakes, an ideal hyperplane for partitioning would significantly separate from any class's nearest planned information aim. To keep processing costs down, the SVM approach computes the minor items about the first-degree variable [15, 16].

4. Results and Discussion

This study's first stage in data analysis was sorting the data according to its various attributes and categories, like crop kind, yield, condition, etc. The results of our evaluations are detailed below. For these evaluations, we tested the proposed strategy and state-of-the-art methods to determine which could reliably generate correct predictions. The dataset has 2200 objects, with 1760 chosen for training and 440 for testing. The training and testing data have been subjected to every known

Table 1. Training datasets

Crop	Label	True	False	True	False
		Positive	Negative	Negative	Positive
Rice	21	19	3	414	17
Maize	12	18	4	385	26
Chickpea	4	18	2	398	33
Kidney beans	10	17	5	421	10
Pigeon bean	19	13	9	412	19
Moth bean	14	15	7	418	13

algorithm and proposed approach. The training dataset details are given in Table 1.

As shown in the table above, the proposed methodology's confusion matrix parameter was examined separately. Whenever something goes wrong that nobody saw coming, this is an example of a false positive. It is determined that the cases are considered to be TN when the expected outcome is negative and the actual result is also harmful. When a predicted negative consequence turns out to be positive, this is an example of a false negative. In addition to comparing the projected algorithm's parameters to those of the standing approaches, a numerical and graphical evaluation of the proposed algorithm's accuracy compared to that of the existing methods should be displayed. All of the graphs show that the suggested strategy has the most value in terms of accuracy, F-score, and recall because it has the highest value. It also has the lowest relative value in terms of both false positives and false negatives. If we are talking about accuracy, the hybrid strategy combining LSTM and RNN is far superior to any other approach. The confusion matrix of positive predicted values of various algorithm are presented in Figure 1.

Accurate prediction of crop yield is crucial for agricultural planning and food security. This paper proposes a novel hybrid method that combines Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) to enhance the precision of crop yield predictions. The method leverages the strengths of both models in handling sequential data and capturing long-term dependencies. Experimental results on various crop datasets demonstrate the hybrid model's superiority over traditional predictive models. While the hybrid model showed improved performance, there are areas for further research. Future work could explore incorporating additional data sources, such as satellite imagery and real-time weather forecasts, to enhance model accuracy. Moreover, the model's applicability to different crops and regions warrants further investigation to ensure its generalizability. The hybrid LSTM-RNN model for crop yield prediction offers several advantages. By combining the strengths of LSTM and RNN layers, it effectively captures both long-term dependencies and short-term sequential patterns, leading to enhanced accuracy. The model's robustness

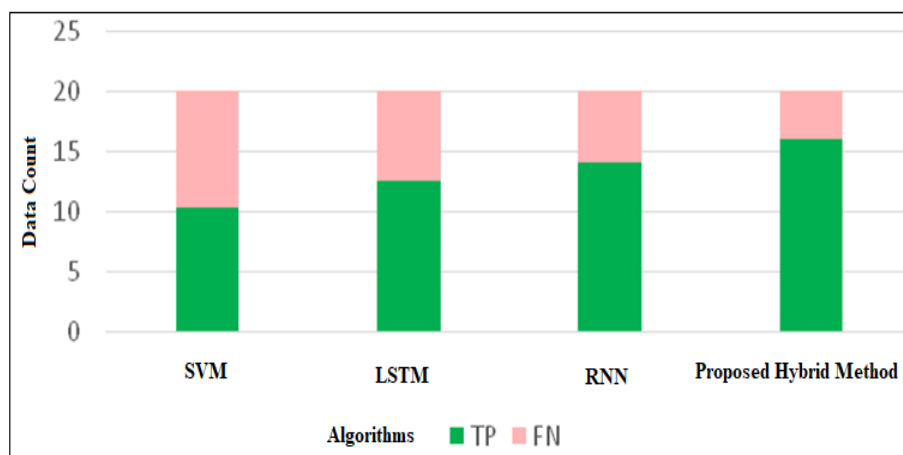


Figure 1. Confusion matrices

allows it to adapt to diverse agricultural data, handle noise, and process large datasets efficiently, making it scalable and suitable for real-world applications.

5. Conclusion

A state-organized dataset of various crops will be utilized in this study to introduce a new hybrid method for agricultural yield estimation. Three components make up the technique that has been suggested: LSTM and RNN combination. The database is responsible for carrying out various preparatory actions to standardize the data gathered. The data is then utilized to train and evaluate both the approaches that are now in use and those that are planned. This approach can achieve an astoundingly high accuracy percentage of 93.02%.

References

- [1] Nguyen-Thanh Son, Chi-Farn Chen, Youg-Sin Cheng, Piero Toscano, Cheng-Ru Chen, Shu-Ling Chen, Kuo-Hsin Tseng, Chien-Hui Syu, Horng-Yuh Guo, and Yi-Ting Zhang. Field-scale rice yield prediction from sentinel-2 monthly image composites using machine learning algorithms. *Ecological informatics*, 69:101618, 2022.
- [2] Venkata Krishna Chaithanya Manam. *Efficient disambiguation of task instructions in crowdsourcing*. PhD thesis, Purdue University Graduate School, 2023.
- [3] Melissa A Stine and Ray R Weil. The relationship between soil quality and crop productivity across three tillage systems in south central honduras. *American Journal of Alternative Agriculture*, 17(1):2–8, 2002.

- [4] Sonal Jain and Dharavath Ramesh. Machine learning convergence for weather based crop selection. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pages 1–6. IEEE, 2020.
- [5] Petteri Nevavuori, Nathaniel Narra, and Tarmo Lipping. Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture*, 163:104859, 2019.
- [6] Ranjini B Guruprasad, Kumar Saurav, and Sukanya Randhawa. Machine learning methodologies for paddy yield estimation in india: a case study. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 7254–7257. IEEE, 2019.
- [7] Patrick Filippi, Edward J Jones, Niranjana S Wimalathunge, Pallegedara DSN Somarathna, Liana E Pozza, Sabastine U Ugbafe, Thomas G Jephcott, Stacey E Paterson, Brett M Whelan, and Thomas FA Bishop. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, 20:1015–1029, 2019.
- [8] VK Chaithanya Manam, Joseph Divyan Thomas, and Alexander J Quinn. Tasklint: Automated detection of ambiguities in task instructions. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 10, pages 160–172, 2022.
- [9] Natarajan Deepa and Kaliyaperumal Ganesan. Decision-making tool for crop selection for agriculture development. *Neural Computing and Applications*, 31:1215–1225, 2019.
- [10] Sanjeev Kulkarni, Aishwarya Shetty, Mimitha Shetty, HS Archana, and B Swathi. Gas spilling recognition and prevention using iot with alert system to improve the quality service. *Perspectives in Communication, Embedded-systems and Signal-processing-PiCES*, 4(4):34–38, 2020.
- [11] Kodimalar Palanivel and Chellammal Surianarayanan. An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, 10(3):110–118, 2019.
- [12] V Suresh Kumar, Sanjeev Kulkarni, Naveen Mukkapati, Abhinav Singhal, Mohit Tiwari, and D Stalin David. Investigation on constraints and recommended context aware elicitation for iot runtime workflow. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3s):96–105, 2024.
- [13] M Kalimuthu, P Vaishnavi, and M Kishore. Crop prediction using machine learning. In *2020 third international conference on smart systems and inventive technology (ICSSIT)*, pages 926–932. IEEE, 2020.
- [14] V K. Chaithanya Manam, Dwarakanath Jampani, Mariam Zaim, Meng-Han Wu, and Alexander J. Quinn. Taskmate: A mechanism to improve the quality of instructions in crowdsourcing. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1121–1130, 2019.
- [15] Kiran Kumar Gopathoti, Anandbabu Gopatoti, Nimma Swathi, and Shamili Srimani Pendyala. Enhancing crop water management: A logistic regression approach integrated with iot for smart irrigation. *International Journal of Scientific Methods in Computational Science and Engineering*, 1(1):1–8, 2024.
- [16] Sanjeev Kulkarni, Sachidanand S Joshi, AM Sankpal, and RR Mudholkar. Link stability based multipath video transmission over manet. *International Journal of Distributed and Parallel Systems*, 3(2):133, 2012.