

Review of: "Adoption of Machine Learning Methods for Crop Yield Prediction-based Smart Agriculture and Sustainable Growth of Crop Yield Production – Case Study in Jordan"

Kiran Raj Bukkarapu

Potential competing interests: No potential competing interests to declare.

The authors have developed a machine learning-based approach to predict crop yields, which is crucial for ensuring global food security and supporting agricultural decision-making. They tested nine different machine learning algorithms. Their integrated model aims to improve the accuracy and reliability of crop yield predictions, helping policymakers with import-export decisions and agricultural planning. However, the manuscript needs significant changes to present their efforts better, which include

1. The title could be refined to enhance clarity and better convey the scope of the study.
2. Consider adding a section to explicitly highlight the "Novelty of the Present Study" and how it advances the field.
3. The sentence, "Machine learning (ML) models are quite tedious because the models of ML for agriculture-based are complex," lacks clarity. Consider revising it for improved precision and flow.
4. The statement "XGBoost, multiple linear regression, Random forest, and Lasso regression give low mean squared errors of 0.092, 0.024, 0.023, and 0.023" lacks units. Including units would aid in understanding these results.
5. Please ensure that Figure 2 includes appropriate units for clarity.
6. Adding a figure that illustrates the food security challenge, sourced from relevant literature, would strengthen the manuscript's context.
7. While the manuscript shows good accuracy, it lacks key details, such as how cross-validation was performed, how tuning parameters were chosen, and how model generalization was assessed. Addressing these would strengthen the study.
8. The manuscript lacks details on hyperparameter tuning, which is crucial for optimizing machine learning models. Please include information on how this was done.
9. It would be helpful to explain how the hyperparameters of the SVM algorithm were optimized. Was a grid search, Bayesian optimization, or another method used?
10. Details such as the split size, number of trees, maximum depth, and features in the random forest model are missing. How were these parameters optimized?
11. Refer to the following manuscripts for additional guidance on model development and tuning:
<https://doi.org/10.1016/j.renene.2022.02.124>, <https://doi.org/10.1177/09544070231158240>,
<https://doi.org/10.1177/14680874241227540>

