

Review of: "Do SHapley Additive exPlanations Enhance Classification Accuracy? An Investigation of Game Theory for Mapping Cityscape Blue-Green Infrastructures Using Sentinel-2"

I. U. Ekanayake¹

¹ University of Peradeniya

Potential competing interests: No potential competing interests to declare.

Good and timely research in applied multidisciplinary research.

Please check on these concerns;

1. Why apply PCA when you have 50K+ datapoints with only 10 variables? What is the purpose of reducing the features?
2. What is the rationale behind employing a random forest classifier? Since the dataset shows high correlation, have you considered using a decision tree or other less complex classification methods?
3. Please add the references for the random forest classifier you have employed (e.g., if it is from scikit-learn, you can find the reference on the webpage)
4. The work you have done is interesting and important, but the title of the paper misleads the readers. As I understood it, SHAP has only been utilized to explain the results here and to get the highest variation of feature importance as the best band; I haven't seen any feedback loop from SHAPly explanations to the random forest classifier. Therefore, I was wondering how the SHAP explanations would enhance the accuracy.
5. Since you are using band selection and testing using the same test dataset, I would suggest you to use a separate cross-validation set to select the bands and test the final results on the test set. (You can decide the best splitting ratios; my suggestion is 60:20:20 (training, testing, CV).)
6. Could you be able to compare the results with the previous publications on this same dataset? Here, I have seen random forest has provided the best accuracies. (<https://www.mdpi.com/2072-4292/12/14/2291>)
7. Refer to this paper to formulate the research findings and comparisons (<https://www.sciencedirect.com/science/article/pii/S2214509522001917>). I highly recommend you to create a result comparison table and a training performance table (loss, accuracy, R2...)