Title: Visualizing Random Forest Classification for Breast Cancer Diagnosis

Summary

This poster provides one pipeline to predict the patients with breast cancer are malignant or benign. After feature standardization, the author constructs random forest model to test the model and then uses the PCA to show the prediction performance.

Strengths

This paper clearly clarifies the intention, data, methodology and conclusion part.

Drawbacks

The article lacks in-depth analysis between the two labels and features.

Evaluation on Clarity and quality of writing (grade: 5)

The work successfully addresses the classification problem and also show the clear logic.

Evaluation on Technical Quality (grade 4)

The techniques are completely correct. But some suggestions are listed below,

1.

Add more scientific evaluation metric to compare the two methods such as AUC in binary classification rather than only using accuracy and RMSE, since if the data exists unbalanced, the accuracy tends to be higher, but the overall performance is still not good. It' s better to show the data distribution with regard to the labels.

2. Could you also further explore why only 2 principal components can have a similar result compared with overall features? Through this way, you can find the 2 principal components' coefficients to see if the most important features with the maximum coefficient is same with your feature possessing the maximum feature importance in random forest model so as to verify your conclusion in the first paragraph of discussion part.

Overall rating: 4.5

Confidence on your assessment: 3