

An improved random forest based on the classification accuracy and correlation measurement of decision trees

Problem they are trying to solve / Purpose of method

Traditional Random Forests is an ensemble of decision trees. It is trained on bootstrap samples. Uses majority voting. All this reduces overfitting.

Issues:

1. Low quality trees
2. High correlation between trees
3. Unstable evaluation

Earlier approaches have only addressed one of the issues.

How does it differ from other methods?

It addresses all issues at the same time.

How the method works

1. Keep only high quality trees
2. Removes highly correlated trees
3. Balanced tree selection process

MORE DETAILS MAYBE?