

Abstract

This study presents experimental investigations on supervised ensemble classification for land cover classification. Despite the arrays of classifiers available in machine learning to create an ensemble, knowing and understanding the correct classifier to use for a particular dataset remains a major challenge. The ensemble method increases classification accuracy by consulting experts taking final decision in the classification process. This study generated various land cover maps, using image classification. This is to authenticate the number of classifiers that should be used for creating an ensemble. The study exploits feature selection techniques to create diversity in ensemble classification. Landsat imagery of Kampala (the capital of Uganda, East Africa), AVIRIS hyperspectral dataset of Indian pine of Indiana and Support Vector Machines were used to carry out the investigation. The research reveals that the superiority of different classification approaches employed depends on the datasets used. In addition, the pre-processing stage and the strategy used during the designing phase of each classifier is very essential. The results obtained from the experiments conducted showed that, there is no significant benefit in using many base classifiers for decision making in ensemble classification. The research outcome also reveals how to design better ensemble using feature selection approach for land cover mapping.

The study also reports the experimental comparison of generalized support vector machines, random forests, C4.5, neural network and bagging classifiers for land cover classification of hyperspectral images. These classifiers are among the state-of-the-art supervised machine learning methods for solving complex pattern recognition problems. The pixel purity index was used to obtain the endmembers from the Indiana pine and Washington DC mall hyperspectral image datasets. Generalized reduced gradient optimization algorithm was used to estimate fractional abundance in the image dataset thereafter obtained numeric values for land cover classification. The fractional abundance of each pixel was obtained using the spectral signature values of the endmembers and pixel values of class labels. As the results of the experiments, the classifiers show promising results. Using Indiana pine and Washington DC mall hyperspectral datasets, experimental comparison of all the classifiers' performances reveals that random forests outperforms the other classifiers and it is computational effective.

The study makes a positive contribution to the problem of classifying land cover hyperspectral images by exploring the use of generalized reduced gradient method and five

supervised classifiers. The accuracy comparison of these classifiers is valuable for decision makers to consider tradeoffs in method accuracy versus complexity. The results from the research has attracted nine publications which include, six international and one local conference papers, one published in Computing Research Repository (CoRR), one Journal paper submitted and one Springer book chapter, Abe *et al.*, 2012 obtained a merit award based on the reviewer reports and the score reports of the conference committee members during the conference period.