Tree Species Classification on Hyperspectral Imagery Using Fewer Training Samples

Fei Tong[1], Yun Zhang[1]

[1] Department of Geodesy and Geomatics Engineering, University of New Brunswick,
15 Dineen Drive, Fredericton, NB E3B 5A3, Canada - (ftong, yunzhang)@unb.ca

Keywords: Active Learning, Hyperspectral Image, Random Forest, SuperPCA, Tree Species Classification.

Abstract

The distribution of tree species within the forest holds significant importance for forest management. Since field surveys in the forest are time-consuming and cost-expensive, automatically extracting tree species distribution maps from remote sensing imagery becomes a trend. For tree species classification using hyperspectral imagery, many existing classification methods require a large number of training samples to achieve high classification accuracy. However, the classification accuracy will decrease rapidly if only a few hundred training samples are used. Given the challenges and expenses associated with collecting abundant training samples in the forest, there is a need to explore methods that achieve good classification performance with a limited number of training samples. In this paper, a classification scheme combining SuperPCA and Active Learning (AL) is proposed to improve the tree species classification using a limited number of training samples. SuperPCA is employed to reduce feature dimensions and harness spectral-spatial information within hyperspectral imagery. Active Learning is employed to select informative samples for the training, thus reducing the requirement for training samples. Experiments on a tree species classification data set demonstrate the effectiveness of the proposed classification scheme.

1. Introduction

Various types of remote sensing sensors, such as hyperspectral, multispectral, and LiDAR, have been used to collect imagery for tree species classification (Fassnacht et al., 2016). Hyperspectral imagery is particularly popular because it provides hundreds of spectral bands, allowing for accurate discrimination between tree species. However, the high dimensionality of hyperspectral data leads to issues such as increased data volume, high correlation between bands, and reduced classification accuracy and efficiency (Zhang et al., 2020). Therefore, to achieve high tree species classification accuracy, the first challenge is to effectively and efficiently utilize the information within the imagery. The requirement of sufficient training samples for the tree species classification task is another problem. Considering the difficulty and high time cost of collecting samples in the complex forest environment, only a limited number of training samples can be obtained in real applications. However, many existing classification algorithms do not perform well with a small number of training samples. Thus, reducing the need for extensive training data without compromising classification accuracy is the second challenge.

To optimize hyperspectal classification, some studies focus on feature selection and feature reduction (Kumar et al., 2020). For example, a parametric and supervised model called projection pursuit was developed to extract features from hyperspectral imagery (Jimenez and Landgrebe, 1999). In (Archibald and Fann, 2007), an embedded feature-selection algorithm cooperated with support vector machines (SVMs) to perform band selection and classification simultaneously. Then (Pal and Foody, 2010) investigated the effectiveness of different band selection strategies for the hyperspectral image classification task. In (Fassnacht et al., 2014), three feature selection methods were examined with the airborne hyperspectral data for the tree species classification task. Recently, (Likó et al., 2022) investigated the effectiveness of three commonly used dimension reduction methods (Principal Component Analysis (PCA), Minimum Noise Fraction (MNF), and Independent Component Analysis (ICA)) for identifying the tree species structure of a floodplain forest area using a hyperspectral image. Although feature selection and feature reduction methods can decrease the dimensionality of the hyperspectral data, and extract the most discriminative bands or features, it remains challenging to find a universally applicable method for all types of hyperspectral data (Fassnacht et al., 2016).

Another way to optimize hyperspectral classification is to exploit spectral-spatial information within the imagery. Typically, this information is utilized based on local regions, in which pixels have similar spectral characteristics (Fauvel et al., 2013). A simple way to utilize spectral-spatial information is to optimize the pixel-based classification result based on local regions. For instance, (Tarabalka et al., 2009) optimized the pixel-wise classification result by conducting majority voting within segmented local regions. Some studies also conducted classifications based on the information in local regions. To utilize spectral-spatial information, (Pu et al., 2014) conducted the classification according to the similarity calculated from local image patches, and (Chen et al., 2011) conducted the classification based on the joint sparse representation within local image patches. More recently, a two dimensional-CNN (2D-CNN) model was proposed to exploit spectral-spatial information within local image patches (Yue et al., 2015). Following this, several three dimensional-CNN (3D-CNN) models were proposed to utilize spectral-spatial information within the 3D hyperspectral cube (Chen et al., 2016; Li et al., 2017; Paoletti et al., 2018; Hamida et al., 2018). Besides fixed-size image patches, shape-adaptive superpixels are also widely used to exploit spectral-spatial information. For instance, (Fang et al., 2015) adopted over-segmented superpixels in the classification based on sparse representation. Subsequently, (Tong et al., 2017) and (Tong and Zhang, 2021) proposed classification methods that consider information within both fixed-size patches and shape-adaptive superpixels.

In recent years, various methods have been proposed for tree species classification using hyperspectral imagery. (Zhang et al., 2020) introduced a 3D-CNN network to exploit the spectral-spatial information within hyperspectral data. Trained with 2.5% of the labeled samples, the 3D-CNN model achieved a classification accuracy of 95.74%. Following this, (Tong and Zhang, 2022) developed a cascaded multilayer random forest model to classify tree species in the same hyperspectral data, achieving a 97.50% classification accuracy with 2.5% of labeled samples for training. Additionally, a spatial-logical aggregation network with morphological transformation for tree species classification was proposed in (Zhang et al., 2023) and a texture-aware self-attention model was proposed in (Li et al., 2024) for the tree species classification task. Although these methods achieved promising classification performance, the premise was sufficient labeled training samples. For example, as reported by (Tong and Zhang, 2022), the classification accuracy of 3D-CNN dropped to 75.15% when only 0.2% samples were used for training. Therefore, it is important to develop a classification method that requires fewer training samples to achieve satisfactory tree species classification performance.

Active learning (AL) is an effective method for reducing training samples in the classification problem. AL selects the most informative samples that are helpful in finding separating surfaces for different classes (Joshi et al., 2009). As a result, the demand for training samples is effectively reduced. In (Li et al., 2010), the AL based on the entropy of samples was proposed to improve the estimation of the class distributions for the semi-supervised multinomial logistic regression model. Then (Li et al., 2011) introduced a Bayesian approach utilizing the Breaking Tie (BT) AL strategy for hyperspectral image classification. Subsequently, (Sun et al., 2015) adopted AL based on the Markov random field (MRF) regularization to select the most uncertain samples for hyperspectral image classification. Recently, deep learning has been combined with AL to enhance hyperspectral imagery classification. For instance, (Haut et al., 2018) combined the Bayesian-Convolutional Neural Networks with six AL criteria for hyperspectral image classification. Moreover, (Cao et al., 2020) combined the CNN with the AL based on the BT strategy to optimize the classification performance with a minimal number of training samples.

Although AL is effective in reducing the demand for training samples, few studies have applied AL to tree species classification. In this study, we propose a classification scheme based on AL for this task. To reduce feature dimensions and fully utilize the imagery information, we adopt SuperPCA, which conducts general PCA (Vidal et al., 2016) locally on segmented superpixels as described in (Jiang et al., 2018), to extract features for classification. To minimize the need for training samples, we implement an AL process using the BT strategy to select the most informative samples. The effectiveness of the proposed classification scheme is validated on a publicly available hyperspectral tree species classification dataset (Zhang et al., 2020).

2. Methodology

In this paper, a classification scheme combining SuperPCA and Active Learning (AL) is proposed to improve the tree species classification using a limited number of training samples. SuperPCA is employed to reduce feature dimensions and harness spectral-spatial information within hyperspectral imagery. After feature extraction, the Random Forest (RF) (Breiman, 2001) classifier is applied to train the classification model using the initial training set with a few samples. The trained model is then used for predictions on the unlabeled testing set. AL with the Breaking Tie strategy (Luo et al., 2005) is subsequently employed, leveraging the predicted class probability vector to select the most informative samples from the unlabeled testing set. These selected samples are then labeled through expert labeling or field surveys and incorporated into the training set. This updated training set undergoes another round of classification, and this iterative process of classification and AL continues in multiple rounds to refine the classification model. The flowchart of the proposed classification scheme is shown in Figure 1. The details about SuperPCA and AL are introduced in the following sections.

2.1 SuperPCA

SuperPCA is a superpixel-wise PCA approach for dimension reduction and feature extraction (Jiang et al., 2018). Rather than conducting the general PCA approach on the entire image, SuperPCA conducts PCA on every individual segmented local superpixel to make use of the spatial context information in the imagery. Since SuperPCA requires segmented superpixels, it adopts the widely used graph-based ERS segmentation method (Liu et al., 2011) to generate homogeneous regions. In ERS, the input image is mapped to a graph $G = (V, E)$ initially, where V represents the vertex set and E represents the edge set. The segmentation task is then transformed into finding a subset of edges $A \subseteq E$ to divide G into multiple smaller connected subgraphs. Since the goal of the ERS segmentation is to generate compact, homogeneous, and size-balanced superpixels, the following objective function is adopted to determine the A:

$$\max_{A} \{H(A) + \alpha B(A)\}, \qquad \text{s.t. } A \subseteq E \tag{1}$$

In the objective function, $H(A)$ represents the entropy rate, and $B(A)$ represents the balancing function. The coefficient α balances the contribution of $H(A)$ and $B(A)$. The greed algorithm in (Nemhauser et al., 1978) is proved to be effective for finding the optimize solution for this objective function.

Since hyperspectral images contain hundreds of spectral bands, which is too many for ERS segmentation, PCA is first used to reduce the number of bands. The first PCA component, containing most of the image information, is then selected as the input for ERS segmentation. Additionally, the expected number of superpixels N must be specified. Following segmentation, PCA will be conducted based on the information within each superpixel. For a specific superpixel S_i, assuming that the original number of features is L, the expected number of reduced features is D, and the number of pixels in S_i is n_i, after PCA, the original feature space $n_i \times L$ in S_i is transformed to $n_i \times D$. Once PCA on all superpixels is finished, the number of features for all pixels decreases from L to the predetermined parameter D. The subsequent classification will be conducted based on the features extracted by SuperPCA.

2.2 Active Learning

With the extracted features from SuperPCA, an iterative classification process based on AL is conducted to reduce the number of training samples and improve classification accuracy. Assuming that the number of initial training samples is M and the number of classes for the classification is C, the extracted features for the training set $M \times D$ are fed into the commonly used RF classifier. After training, predictions are conducted on the

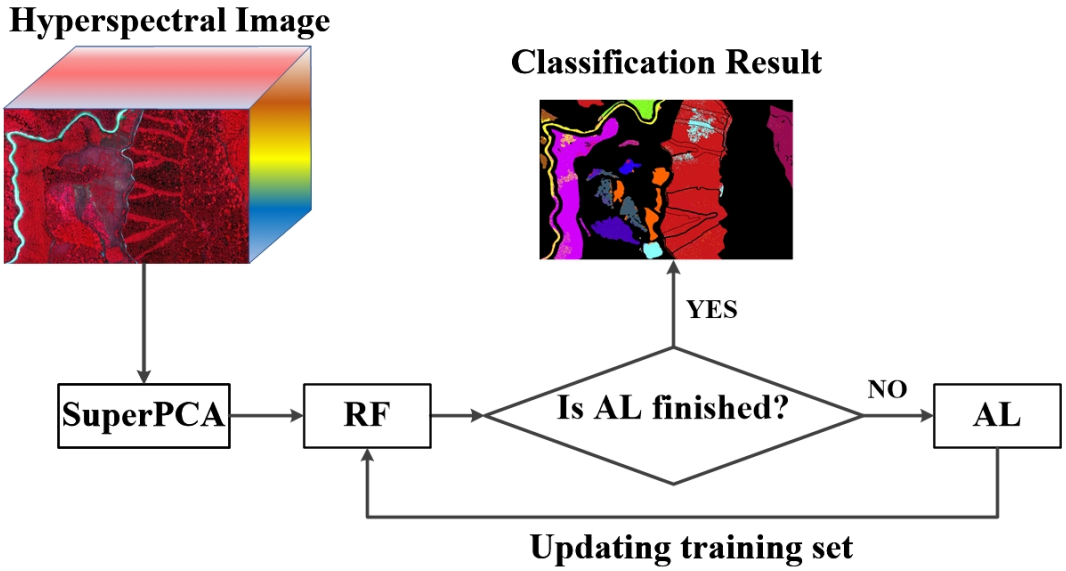


Figure 1. Flowchart of the proposed classification scheme.

testing set. Since the initial number of samples in the training set is few, the predicted results may not be accurate. However, the predicted results are helpful for finding the most informative samples, which are effective for improving the classification model.

For arbitrary pixels in the testing samples, the output for the prediction is the class probability vector $V_i = \{P_i^1, P_i^2, \ldots, P_i^C\}$. Then the AL based on the BT strategy (Luo et al., 2005) is adopted to select the most informative samples from the testing set. In the BT strategy, the uncertainty of samples in the multiclass classification problem is evaluated, with the most uncertain samples considered the most informative. The uncertainty in the BT strategy is calculated from the top two probability values in the class probability vector. This approach ignores classes with low probabilities, avoiding the negative influence of these less important classes. For arbitrary class probability vector $V_i = \{P_i^1, P_i^2, \ldots, P_i^C\}$, the uncertainty U_i is calculated as:

$$U_i = P_i^A - P_i^B \tag{2}$$

where A represents the class with the greatest class probability value in V_i, and B represents the class with the second greatest class probability value in V_i. If the difference between P_i^A and P_i^B is small, it indicates that the trained classification model struggles to distinguish between classes A and B. Therefore, samples in the testing set with small U values are considered as informative. In each round of AL, K informative samples are selected from the unlabelled testing set. Subsequently, these K samples are annotated by expert labeling or field survey. Then these samples are added to the training set. The updated training set is used for a new round of training with the RF classifier. This iterative AL process repeats until the predetermined number of rounds R is reached. After R rounds of AL, $K \times R$ samples are selected from the testing set and added to the training set. The output of the final round of training is the class probability vector for each pixel. The final label of each pixel in the testing set is determined by selecting the class with the highest probability.

3. Experiments

3.1 Dataset

To verify the effectiveness of the proposed classification scheme, experiments were conducted on a publicly available hyperspectral tree species classification dataset (Zhang et al., 2020). The size of the image is 572×906, with a spatial resolution of 1 m. It comprises 125 spectral bands with coverage ranging from 400 nm to 1000 nm. The spectral resolution is 3.3 nm. Labeled samples were collected for 12 classes, comprising 9 forest vegetation classes and 3 non-forest vegetation classes. The number of labeled samples for all classes and the corresponding number of training samples utilized in the proposed classification scheme are listed in Table 1. The hyperspectral image and corresponding ground truth map are shown in Figure 2. The initial number of training samples for each class was 5. The entire classification process involved 10 rounds of AL to update the training set. In each AL round, 10 unlabeled samples selected from the testing set were labeled and incorporated into the training set. Thus, the total number of training samples used in the proposed classification scheme was 160.

3.2 Classification Performance

Classification accuracy of each class, overall accuracy (OA), average accuracy (AA), and Kappa coefficient were adopted to evaluate the performance of the classification. The comparison was conducted with the 3D-CNN model (Zhang et al., 2020) applied to the same dataset. Since 3D-CNN didn't include the AL process, the initial number of training samples for each class was set to 20. The total number of training samples used in 3D-CNN was 240. All hyperparameters for the 3D-CNN model were either set according to (Zhang et al., 2020) or tuned for optimal classification performance on the testing set. Moreover, to highlight the improvement introduced by SuperPCA and AL, two additional classifications were conducted: (1) RF + Original Features + AL, which replaced the extracted SuperPCA features with the original spectral features in hyperspectral data, (2) RF + SuperPCA, which conducted a one-time classification with RF but the initial number of training samples for each class was set as 20. For the parameters used in SuperPCA,

Table 1. Number of samples in the tree species dataset and number of training samples used in the proposed classification scheme.

Class Names	Number of Samples	Initial Training Samples	Training Samples Added by AL
Cunninghamia lanceolata	80743	5	
Pinus massoniana	5340	5	
Pinus elliottii	4175	5	
Eucalyptus grandis x urophylla	14931	5	
Eucalyptus urophylla	40275	5	
Castanopsis hystrix	25461	5	10 AL rounds
Mytilaria laosensis	3000	5	(Each round adding 10 samples)
Camellia oleifera	8324	5	
Other broadleaf forest	10741	5	
Road	10472	5	
Cutting blank	10170	5	
Building land	68	5	
Total Number	213700	60	100

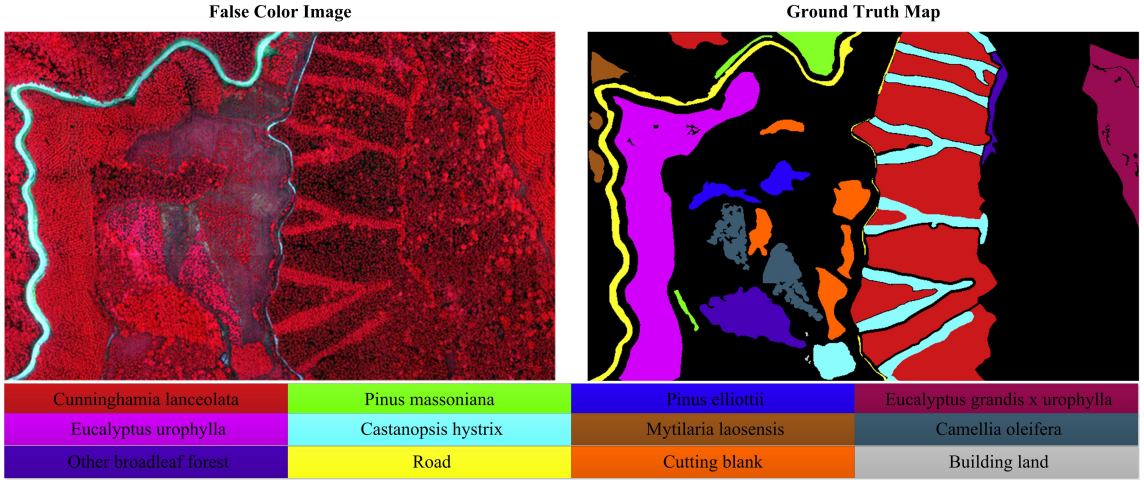


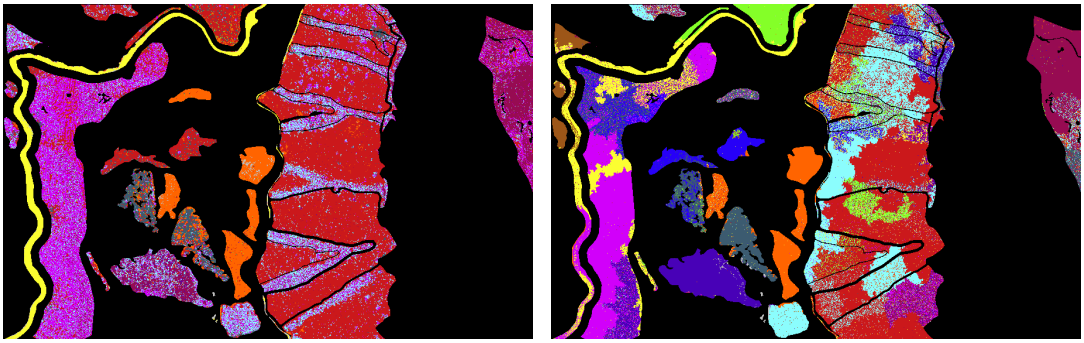
Figure 2. Hyperspectral image (false color) and ground truth map.

Table 2. Classification accuracy of different classification methods.

	RF + Original Features + AL	RF + SuperPCA	3D-CNN	Proposed
Cunninghamia lanceolata	84.61	49.28	82.16	87.17
Pinus massoniana	19.60	88.69	57.76	70.36
Pinus elliottii	5.57	84.61	61.57	62.12
Eucalyptus grandis x urophylla	47.80	93.82	66.65	91.03
Eucalyptus urophylla	70.47	66.59	56.05	91.02
Castanopsis hystrix	24.00	49.27	51.59	36.06
Mytilaria laosensis	3.13	95.96	83.67	77.03
Camellia oleifera	51.95	80.16	87.15	74.31
Other broadleaf forest	33.24	85.68	80.34	82.30
Road	88.35	65.77	92.17	52.75
Cutting blank	89.74	92.34	97.12	89.49
Building land	83.96	100	100	100
OA(%)	64.42	63.87	72.83	78.71
AA(%)	50.20	79.35	76.35	76.14
Kappa×100	53.92	57.42	67.06	72.57
Number of Training Samples	160	240	240	160

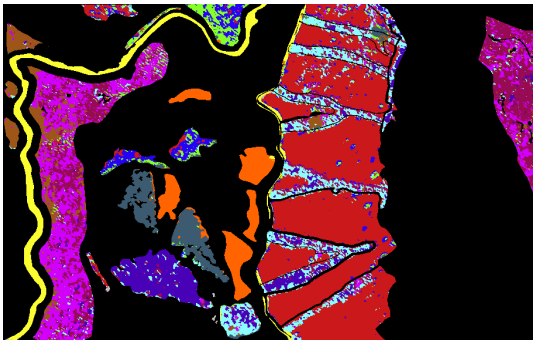
the number of superpixels was set to 100 and the number of reduced features was set to 30. The RF classifier used 500 trees for training. To ensure a fair comparison with non-AL classification methods, all initial training samples used in AL methods were drawn from the 240 labeled samples used in non-AL methods. The evaluation metrics of all classifications are tabulated in Table 2. All the metrics were averaged by ten runs with different random training samples. The classification maps for all methods from one of the ten runs are shown in Figure 3.

As can be observed in Table 2, the proposed method achieved the highest classification accuracy among all methods. Although the proposed classification method used 80 fewer samples than 3D-CNN for training, it achieved an accuracy increase of 5.88%. For the nine forest vegetation classes, the proposed method outperformed the 3D-CNN in six classes. In the result of 3D-CNN, the classification accuracy of five forest vegetation classes was lower than 70%. For the classes "Pinus elliottii" and "Castanopsis hystrix", the accuracy of the proposed method was lower than 70%. This low accuracy could be due to the imbalanced distribution of samples among different classes. Specifically, "Cunninghamia lanceolata" and "Eucalyptus urophylla" constituted 37.78% and 18.85% of the available labeled samples, respectively, while the remaining 10 classes shared 43.42% of the samples. The proposed classification method achieved high classification accuracy (87.17% and 91.02% re-

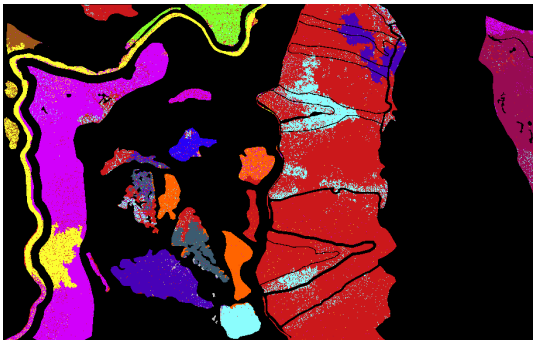


(a) RF + Original Features + AL

(b) RF + SuperPCA



(c) 3D-CNN



(d) Proposed

Figure 3. Tree species classification maps generated from different methods.

spectively) for these two dominant classes. Since these two dominant classes occupy high proportions in available labeled samples, they have a higher probability of being selected in the AL process. Thus, for 100 extra samples selected by AL, these two dominant classes received more training samples than other classes. The accuracy of the proposed method was 14.29% higher than that of the method RF + Original Features + AL, demonstrating that features extracted using SuperPCA significantly outperformed the original spectral features. Moreover, compared with RF + SuperPCA, the proposed method achieved 14.84% higher classification accuracy with fewer training samples, proving that AL is beneficial for improving classification accuracy and reducing the number of required training samples.

From the classification maps shown in Figure 3, it is evident that the RF + Original Features + AL method produced a lot of pepper noise due to the lack of spatial information in the classification process. In the RF + SuperPCA results, significant misclassifications occurred in the two dominant forest vegetation classes, as 20 samples were insufficient for satisfactory model performance. The 3D-CNN results showed clear confusion between the classes "Eucalyptus grandis x urophylla" and "Eucalyptus urophylla". In contrast, the classification map generated by the proposed method exhibited fewer misclassifications compared to other methods.

4. Conclusion

This paper proposes an effective classification scheme to conduct tree species classification using a few training samples. Using 160 training samples, our proposed classification scheme achieved a classification accuracy of 78.71%. Notably, the accuracy of the proposed classification scheme is 5.88% higher than that of the 3D-CNN. Moreover, compared to the other two classifications, our scheme demonstrates a remarkable accuracy increase of over 14%, highlighting the effectiveness of SuperPCA and AL in enhancing tree species classification accuracy. It's worth mentioning that the RF classifier utilized in our proposed scheme is interchangeable with other classifiers. Therefore, employing a more sophisticated classifier could potentially further enhance the classification performance.

References

Archibald, R., Fann, G., 2007. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geosci. Remote Sens. Lett.*, 4(4), 674–677.

Breiman, L., 2001. Random Forests. *Mach. Learn.*, 45(1), 5–32.

Cao, X., Yao, J., Xu, Z., Meng, D., 2020. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans. Geosci. Remote Sens.*, 58(7), 4604–4616.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.*, 54(10), 6232-6251.

Chen, Y., Nasrabadi, N. M., Tran, T. D., 2011. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.*, 49(10), 3973–3985.

Fang, L., Li, S., Kang, X., Benediktsson, J. A., 2015. Spectral–Spatial Classification of Hyperspectral Images With a Superpixel-Based Discriminative Sparse Model. *IEEE Trans. Geosci. Remote Sens.*, 53(8), 4186–4201.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.*, 186, 64–87.

Fassnacht, F. E., Neumann, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P. K., Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, 7(6), 2547–2561.

Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., Tilton, J. C., 2013. Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proc. IEEE*, 101(3), 652–675.

Hamida, A. B., Benoit, A., Lambert, P., Amar, C. B., 2018. 3-D deep learning approach for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.*, 56(8), 4420–4434.

Haut, J. M., Paoletti, M. E., Plaza, J., Li, J., Plaza, A., 2018. Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. *IEEE Trans. Geosci. Remote Sens.*, 56(11), 6440–6461.

Jiang, J., Ma, J., Chen, C., Wang, Z., Cai, Z., Wang, L., 2018. SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.*, 56(8), 4581-4593.

Jimenez, L. O., Landgrebe, D. A., 1999. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Trans. Geosci. Remote Sens.*, 37(6), 2653–2667.

Joshi, A. J., Porikli, F., Papanikolopoulos, N., 2009. Multi-class active learning for image classification. *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2372–2379.

Kumar, B., Dikshit, O., Gupta, A., Singh, M. K., 2020. Feature extraction for hyperspectral image classification: A review. *Int. J. Remote Sens.*, 41(16), 6248–6287.

Li, J., Bioucas-Dias, J. M., Plaza, A., 2010. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sens.*, 48(11), 4085–4098.

Li, J., Bioucas-Dias, J. M., Plaza, A., 2011. Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Trans. Geosci. Remote Sens.*, 49(10), 3947–3960.

Li, N., Jiang, S., Xue, J., Ye, S., Jia, S., 2024. Texture-Aware Self-Attention Model for Hyperspectral Tree Species Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-15.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.*, 9(1), 67.

Likó, S. B., Bekő, L., Burai, P., Holb, I. J., Szabó, S., 2022. Tree species composition mapping with dimension reduction and post-classification using very high-resolution hyperspectral imaging. *Scientific Reports*, 12(1), 20919.

Liu, M., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation. *CVPR 2011*, 2097–2104.

Luo, T., Kramer, K., Goldgof, D. B., Hall, L. O., Samson, S., Remsen, A., Hopkins, T., Cohn, D., 2005. Active learning to recognize multiple types of plankton. *J. Mach. Learn. Res.*, 6(4).

Nemhauser, G. L., Wolsey, L. A., Fisher, M. L., 1978. An analysis of approximations for maximizing submodular set functionsI. *Math. Program.*, 14, 265–294.

Pal, M., Foody, G. M., 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.*, 48(5), 2297–2307.

Paoletti, M. E., Haut, J. M., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.*, 145, 120–147.

Pu, H., Chen, Z., Wang, B., Jiang, G. M., 2014. A Novel Spatial–Spectral Similarity Measure for Dimensionality Reduction and Classification of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.*, 52(11), 7008–7022.

Sun, S., Zhong, P., Xiao, H., Wang, R., 2015. An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery. *IEEE J. Sel. Top. Signal Process.*, 9(6), 1074–1088.

Tarabalka, Y., Benediktsson, J. A., Chanussot, J., 2009. Spectral–Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques. *IEEE Trans. Geosci. Remote Sens.*, 47(8), 2973–2987.

Tong, F., Tong, H., Jiang, J., Zhang, Y., 2017. Multiscale Union Regions Adaptive Sparse Representation for Hyperspectral Image Classification. *Remote Sens.*, 9(9).

Tong, F., Zhang, Y., 2021. Exploiting Spectral–Spatial Information Using Deep Random Forest for Hyperspectral Imagery Classification. *IEEE Geosci. Remote Sens. Lett.*, 19, 1–5.

Tong, F., Zhang, Y., 2022. Spectral–spatial and cascaded multilayer random forests for tree species classification in airborne hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–11.

Vidal, R., Ma, Y., Sastry, S. S., 2016. *Principal Component Analysis*. Springer New York, New York, NY, USA, 25–62.

Yue, J., Zhao, W., Mao, S., Liu, H., 2015. Spectralspatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sens. Lett.*, 6(6), 468-477. https://doi.org/10.1080/2150704X.2015.1047045.

Zhang, B., Zhao, L., Zhang, X., 2020. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.*, 247, 111938.

Zhang, M., Li, W., Zhao, X., Liu, H., Tao, R., Du, Q., 2023. Morphological transformation and spatial-logical aggregation for tree species classification using hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 61, 1–12.