environmental sciences proceedings

MDPI

Proceedings

Comparison of supervised classification algorithms using a hyperspectral image for land use land cover classification

Sonia Sharma Banjade*, Bipana Subedi and Nitant Rai

Virginia Polytechnic Institute and State University, Forest Resources and Environmental Conservation, Blacksburg; VA 24060, USA;
soniasharma07@vt.edu (S.S.B.); sbipana98@vt.edu (B.S.); nitant3@vt.edu (N.R.)
* Correspondence: soniasharma07@vt.edu

Abstract: Hyperspectral Imaging is getting popular in land use land classification because of its ability to capture detailed information through higher spatial resolution and contagious spectral bands. Using the hyperspectral image from G-LiHT (Goddard's LiDAR, Hyperspectral, and Thermal) Airborne Imager covering a study area in Tennessee, Knoxville, we compared the performance of Spectral Angle Mappers (SAM), Spectral Information Divergence (SID), and Support Vector Machine (SVM) for land use land cover classification. We used a confusion matrix for the accuracy assessment of the classifiers. Among the three classifiers, SVM showed the highest accuracy with 92.03%. Our results also show that some classes, such as water and forests, are consistently distinguishable across all classification methods, while others, such as built-up areas vary depending on the technique used.

Keywords: supervised classification; hyperspectral image; land use; spatial resolution; classifiers

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date

1. Introduction

The detailed mapping of land cover land use change has advanced in recent days with emerging satellite data based on multispectral or hyperspectral sensors. Since the 1960s, remote sensing data has been used in land cover mapping. The detailed mapping assists in analyzing changes over time in various land use land cover classes and assessing risk at various scales [1]. This information plays a vital role in preserving ecologically sensitive areas and solving environmental issues. With the increasing urbanization and land degradation, the significance of land use classification has increased [2]

A wide range of multispectral images are mostly used in image classification. However, with the rapid advancement in technology, hyperspectral images are also being used in recent days. The hyperspectral images provide spectral data for each pixel in numerous contiguous spectral bands often covering a wide range of wavelengths [3] Also, these images have high spectral resolution making them able to capture detailed information about the spectral characteristics of the observed objects or surface.

The development of reliable image classification depends on the performance of classification algorithms. Specifically dealing with hyperspectral images, the high dimensionality and spectral mixing are the major challenges [4] which can significantly impact the accuracy of classification results. Additionally, an inadequate number of ground truth data, as well as potential redundancy in hyperspectral images, add complexity to the classification process [5]. Therefore, this study aims to investigate the robustness of classification algorithms in handling spectral unmixing and limited ground truth information. compare the various image classification algorithms of a hyperspectral image.

An image classification involves a process where each individual pixel within the image is categorized into discrete land use classes [6,7]. The most often used two methods of classification are supervised classification and unsupervised classification. In supervised classification, the analyst knows about the labels of classes or has the training data set [8] before implying the classification in the image.

Hence this study aims to perform and compare supervised classifiers (SVM, SAM, and SID) for the land use land cover classification using hyperspectral image.

2. Materials and Methods

2.1. Data

The hyperspectral image from the G-LiHT (Goddard's LiDAR, Hyperspectral, and Thermal) was used for this research. This data has undergone processing to generate standardized data products, including 1-meter at-sensor reflectance hyperspectral imagery. The flight was conducted on May 7, 2015, in Stanton of Knoxville, Tennessee. The image was acquired as UTK_7May2015_Stanton and was downloaded from the NASA G-LiHT website (https://glihtdata.gsfc.nasa.gov/) with 119 spectral bands between 418 and 918 nm. The hyperspectral imaging spectrometer model was Hyperspec model 1002A-00451; Headwall Photonics [9].

2.2. Data pre-processing, training, and testing dataset

We preprocessed the hyperspectral image to reduce redundancy and noise. We used Minimum Noise Fraction (MNF) transformation, a widely-used technique that serves to de-correlate spectral bands and reduce noise, effectively isolating the signal from undesirable variations [10]. Following this transformation, an eigenvalue analysis is conducted to determine the importance of each MNF component. MNF components with higher eigenvalues are prioritized as they capture more detailed information about the land cover classes being analyzed. We subsetted the original image with 30 bands only reducing the band with low eigenvalues.

We used the spectral library created through an in-field survey using a spectrometer as the ground truth data or training data set for image classification. We used a random sampling method to choose our samples in the field. A total of 60 samples of each land cover type (vegetation, grassland, built-up, bare soil, water) were recorded using a spectrometer. Then, 40 samples were chosen for the training dataset randomly. The remaining 20 samples were chosen for testing.

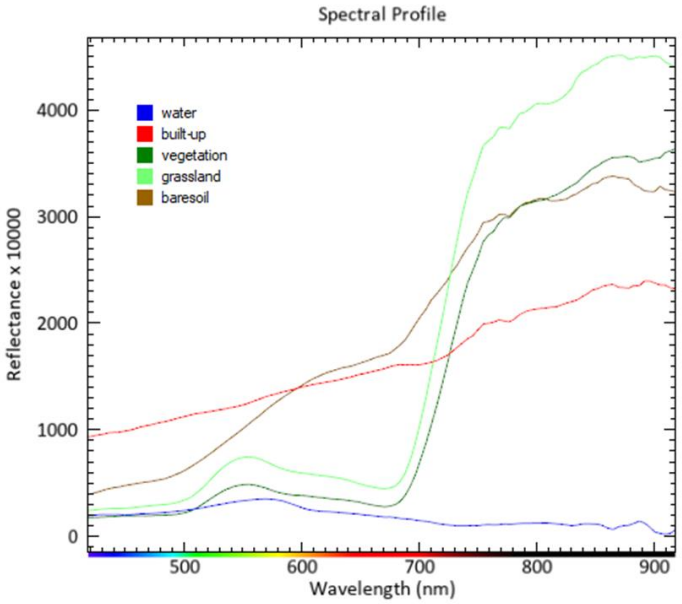

Figure 1. Spectral library of different landcover types.

2.3. Image classification

2.3.1. Spectral Angle Mappers (SAM)

SAM is particularly valuable for identifying and characterizing materials or objects within a scene based on their spectral signatures. SAM operates on the principle that the similarity between two spectra can be quantified by measuring the angle between them in a high-dimensional space, where each dimension corresponds to a spectral band or wavelength. [11] For this classification method, we tried various values of maximum angle radians, and the best result was obtained when a value of 0.3 was used.

2.3.2 Spectral Information Divergence (SID)

SID is commonly used in hyperspectral image processing for tasks like anomaly detection, target detection, and classification. It helps identify areas or objects in an image that deviate significantly from the expected spectral distribution, which can be useful in image classification [11] For the SID algorithm, we used the maximum divergence threshold's value of 0.5 to obtain the best result.

2.3.3 Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used for classification and is effective in high-dimensional images [12]. We used radial radial-based kernel function using a gamma value of 0.009 for SVM.

2.4. Accuracy assessment

The accuracy assessment is a crucial step in land use land cover classification for the validation of the classified image. We used a confusion matrix, which summarizes the class labels against the predicted labels to evaluate the performance of supervised classification algorithms. The total accuracy was calculated as:

Overall accuracy = (Number of correctly classified pixels ÷ Total number of pixels) * 100

3. Results

3.1. Image Classification

SAM appears to give no data value surrounding the water bodies and built-up areas. SID is able to remove the no data value from the image. It can be seen that the forest and built-up areas are clearly classified. Although, there seems to be some noise over the water bodies. While SVM performed well in detecting the land cover types removing the no data value over the water bodies and surrounding built-up areas.

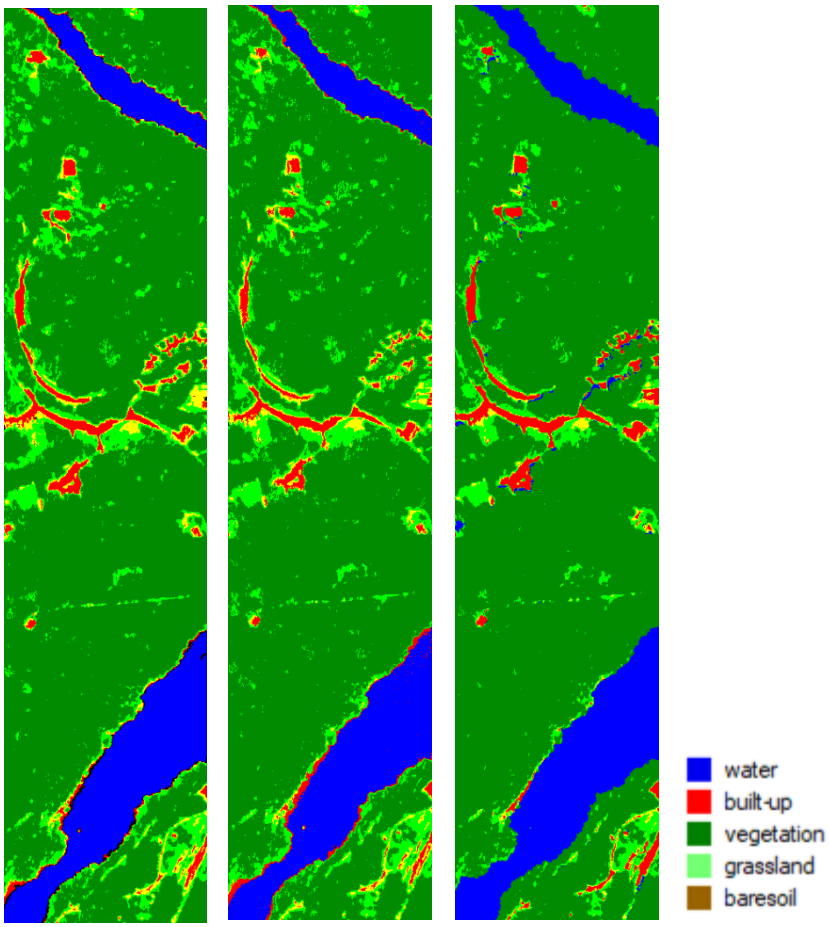

Figure 2. Supervised classification using SAM, SID, and SVM (left to right) showing five classes and unclassified labels.

3.2. Accuracy assessment

The confusion matrix table for each of the classification algorithms (SVM, SID, and SAM) is represented in Table 1, Table 2, and Table 3.

SVM achieved an exceptional accuracy of 92.03% , SID had 89.60% and SAM had 91.23%. The confusion matrices provide further insights into the classification performance, detailing the distribution of true positives (correctly classified pixels), true negatives, false positives, and false negatives for each classifier. The high values in the diagonal of the confusion matrices indicate strong agreement between predicted and actual class labels.

Table 1. Accuracy assessment of SVM.

Class	Baresoil	Grassland	Water	Built-up	Vegetation	Total
Unclassified	0	0	0	0	0	0
Baresoil	77.62	2.98	0	2.46	0	1.31
Grassland	14.69	94.47	0	7.91	1.09	3.71
Water	0	0	99.25	4.61	0	23.68
Built-up	6.99	0	0.75	41.55	0.08	5.06
Vegetation	0.7	2.55	0	43.47	98.83	66.24
Total	100	100	100	100	100	100

Overall Accuracy = 92.03%

Table 2. Accuracy assessment of SID.

Class	Baresoil	Grassland	Water	Built-up	Vegetation	Total
Unclassified	0	0	0	0	0	0
Baresoil	79.02	3.4	0.3	10.91	0.03	2.39
Grassland	20.98	96.6	0	13.52	2.98	5.64
Water	0	0	96.85	0	0	22.6
Built-up	0	0	2.85	34.79	0.03	4.65
Vegetation	0	0	0	40.78	96.96	64.72
Total	100	100	100	100	100	100

Overall Accuracy = 89.60%

Table 3. Accuracy assessment of SAM.

Class	Baresoil	Grassland	Water	Built-up	Vegetation	Total
Unclassified	0	0	1.43	0	0	0.33
Baresoil	92.31	2.98	0.23	8.76	0.08	2.32
Grassland	7.69	97.02	0	10.6	2.08	4.59
Water	0	0	97.52	0	0	22.75
Built-up	0	0	0.83	41.55	0.03	4.95
Vegetation	0	0	0	39.09	97.81	65.05
Total	100	100	100	100	100	100

Overall Accuracy = 91.23%

4. Discussion

The accuracy assessment results of the land use land cover classification, employing SVM, SAM, and SID classifiers, reveal promising outcomes for the hyperspectral image. The achieved accuracies for all three classes indicate they performed really well for detailed classification. While SVM stands as a top-performing classifier with the highest accuracy of 92.03% among the three of them. The result was consistent with a comparative study on the effectiveness of image classification algorithms, including SVM, SAM, and SID conducted by [14] and [15], which also concluded that SVM performs better than other methods. Despite showing the highest accuracy, SVM is computationally intensive, especially with large datasets. Though SAM had negligible differences with SVM, SAM is proven to be best in capturing spectral similarity based on spectral angles [15].

The notable outcomes of this research are the consistency of distinguishability of forest and water across all employed classification schemes. This implies the spectral signatures of these classes are distinct and easily discernible by the selected classifiers. In contrast, variability is seen in built-up areas. Also, the challenge seen in this research is the shadow, particularly tall structures, and trees. In many cases, these shadows create dark pixels within the image and can be incorrectly classified as water bodies.

5. Conclusion

Following an analysis of different supervised image classifiers, we discovered that SVM outperforms other classifiers in accurately identifying land cover land use classes and is also effective at handling high-dimensional data. Following SVM, SAM can also serve as a suitable method for detecting land cover land use classes, as there was negligible difference between SAM and SVM. The detection in built-up areas and water bodies is slightly mislabeled with shadow by SID. Whereas, the SVM demonstrated its effectiveness in handling such scenarios. Hence, these three supervised classifiers were demonstrated to be effective in classifying remotely sensed data.

Author Contributions: Conceptualization, S.S.B.; Data curation, S.S.B., and N.R.; Formal analysis, S.S.B., N.R., and B.S.; Methodology, S.S.B.; Writing – original draft, S.S.B., and B.S.; Writing – review & editing, S.S.B., N.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be available on request.

Acknowledgments: We would like to thank Mr. Santosh Rijal for having the lab sessions in the "Remote Sensing" course where we were able to collect field spectrometer data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Petropoulos, G.P.; Arvanitis, K.; Sigrimis, N. Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Syst. Appl.* **2012**, *39*, 3800–3809, doi:10.1016/j.eswa.2011.09.083.
2. Shafri, H.Z.M.; Suhaili, A.; Mansor, S. The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. *J. of Computer Science* **2007**, *3*, 419–423, doi:10.3844/jcssp.2007.419.423.
3. Raczko, E.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *EuJRS* **2017**, *50*, 144–154, doi:10.1080/22797254.2017.1299557.
4. Deilmai, B.R.; Ahmad, B.B.; Zabihi, H. Comparison of two Classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. *IOP Conf. Ser.: Earth Environ. Sci.* **2014**, *20*, 012052, doi:10.1088/1755-1315/20/1/012052.
5. Ballanti, L.; Blesius, L.; Hines, E.; Kruse, B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sens (Basel)* **2016**, *8*, 445, doi:10.3390/rs8060445.
6. Barnsley, M. J., & Barr, S. L. (1996). Inferring... - Google Scholar. (accessed on 25 September 2023).
7. Barnsley, M.J.; Barr, S.L. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric engineering and remote sensing* **1996**.
8. Bagirov, A.M.; Rubinov, A.M.; Soukhoroukova, N.V.; Yearwood, J. Unsupervised and supervised data classification via nonsmooth and global optimization. *TOP* **2003**, *11*, 1–75, doi:10.1007/BF02578945.
9. Cook, B. D., L. W. Corp, R. F. Nelson, E. M. Middleton, D. C. Morton, J. T. McCorkel, J. G. Masek, K. J. Ranson, and V. Ly. 2013. NASA Goddard's Lidar, Hyperspectral and Thermal (G-LiHT) airborne imager. Remote Sensing 5:4045-4066, doi:10.3390/rs5084045.
10. Luo, G.; Chen, G.; Tian, L.; Qin, K.; Qian, S.-E. Minimum Noise Fraction versus Principal Component Analysis as a Preprocessing Step for Hyperspectral Imagery Denoising. *Canadian Journal of Remote Sensing* **2016**, *42*, 106–116, doi:10.1080/07038992.2016.1160772.
11. de Jr, O.; Meneses, P. Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM).
12. Chein-I Chang Spectral information divergence for hyperspectral image analysis. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*; IEEE, 1999; pp. 509–511.
13. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567, doi:10.1038/nbt1206-1565.
14. Shakya, A.K.; Ramola, A.; Kandwal, A.; Prakash, R. Comparison of supervised classification techniques with alos palsar sensor for roorkee region of uttarakhand, india. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *XLII–5*, 693–701, doi:10.5194/isprs-archives-XLII-5-693-2018.
15. Wang, K.; Cheng, L.; Yong, B. Spectral-Similarity-Based Kernel of SVM for Hyperspectral Image Classification. *Remote Sens (Basel)* **2020**, *12*, 2154, doi:10.3390/rs12132154.
16. Dennison, P.E.; Halligan, K.Q.; Roberts, D.A. A comparison of error metrics and constraints for multiple endmember spectral

mixture analysis and spectral angle mapper. *Remote Sensing of Environment* **2004**, *93*, 359–367, doi:10.1016/j.rse.2004.07.013.

Cook, B. D., L. W. Corp, R. F. Nelson, E. M. Middleton, D. C. Morton, J. T. McCorkel, J. G. Masek, K. J. Ranson, V. Ly, and P. M. Montesano. 2013. NASA Goddard's Lidar, Hyperspectral and Thermal (G-LiHT) airborne imager. Remote Sensing 5:4045-4066, doi:10.3390/rs5084045.